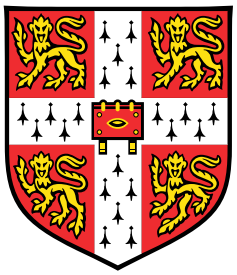
Detailed and Practical 3D Reconstruction with Advanced Photometric Stereo Modelling



Fotios Logothetis

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College October 2018

Detailed and Practical 3D Reconstruction with Advanced Photometric Stereo Modelling

Fotios Logothetis

Object 3D reconstruction has always been one of the main objectives of computer vision. After many decades of research, most techniques are still unsuccessful at recovering high resolution surfaces, especially for objects with limited surface texture. Moreover, most shiny materials are particularly hard to reconstruct.

Photometric Stereo (PS), which operates by capturing multiple images under changing illumination has traditionally been one of the most successful techniques at recovering a large amount of surface details, by exploiting the relationship between shading and local shape. However, using PS has been highly impractical because most approaches are only applicable in a very controlled lab setting and limited to objects experiencing diffuse reflection.

Nevertheless, recent advances in differential modelling have made complicated Photometric Stereo models possible and variational optimisations for these kinds of models show remarkable resilience to real world imperfections such as non-Gaussian noise and other outliers. Thus, a highly accurate, photometric-based reconstruction system is now possible.

The contribution of this thesis is threefold. First of all, the Photometric Stereo model is extended in order to be able to deal with arbitrary ambient lighting. This is a step towards acquisition in a non-fully controlled lab setting. Secondly, the need for a priori knowledge of the light source brightness and attenuation characteristics is relaxed as an alternating optimisation procedure is proposed which is able to estimate these parameters. This extension allows for quick acquisition with inexpensive LEDs that exhibit unpredictable illumination characteristics (flickering etc). Finally, a volumetric parameterisation is proposed which allows one to tackle the multi-view Photometric Stereo problem in a similar manner, in a simple unified differential model. This final extension allows for complete object reconstruction merging information from multiple images taken from multiple viewpoints and variable illumination.

The theoretical work in this thesis is experimentally evaluated in a number of challenging real world experiments, with data captured by custom-made hardware. In addition, the applicability of the generality of the proposed models is demonstrated by presenting a differential model for the shape of polarisation problem, which leads to a unified optimisation problem, fusing information from both methods. This allows for the acquisition of geometrical information about objects such as semi-transparent glass, hitherto hard to deal with.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Fotios Logothetis
October 2018

Acknowledgements

Firstly, I wish to thank my supervisor Roberto Cipolla for his support and advice, which has made working with him a very memorable experience.

I would also like to thank my collaborators: Charalampos Konstantinou for the stone scanning project, Yvain Quéau for the ambient light project and Stefano Bucciarelli for assisting in the making of the camera control code.

My warmest thanks goes to all of my lab colleagues in the Department of Engineering, University of Cambridge. In particular: Benjamin Biggs, Ignas Budvytis, James Charles, Federica Freddi, Ankur Handa, Je Hyeong Hong, Yani Ioannou, Alex Kendall, Viorica Patraucean, Thomas Roddick. A special mention has to go to our amazing IT officer Peter Grandi for just being there to solve our problems.

I would also like to thank my friends Theocharis Amanatidis, Grigorios Gakis, Dimitris Kousoulidis, Andreas Petrides and Agavi Stavropoulou-Tatla, for their constructive advice and invaluable support. I will be forever grateful to them and to Dimitris Bakalis, Alkis Gkouzias and Alex Mitropoulos, for helping me preserve my sanity at times when everything seemed to go wrong.

I am also grateful to my aunt Sophia Travlou for proofreading this dissertation.

I must also express my undying gratitude to my parents and brother for their persistent encouragement and love which has supported me in all my endeavours.

Last, but not least, I have to thank my mentor and friend Roberto Mecca. He introduced me to the Photometric Stereo topic, and ever since has been an invaluable source of support in many ways. Therefore, this thesis is dedicated to him.

Contents

List of Figures

Symbols

Abbreviations

ADMM Alternating Direction Method of Multipliers

BRDF Bidirectional Reflectance Distribution Function

CPU Central Processing Unit

GPU Graphics Processing Unit

LED Light-Emitting Diode

MAE Mean Angular Error

MVPS Multi-view Photometric Stereo

MVS Multi-view Stereo

MVSFS Multi-view Shape from Shading

ODE Ordinary Differential Equation

PDE Partial Differential Equation

PS Photometric Stereo

SDF Signed Distance Function/Field

SFM Structure From Motion

SFP Shape from Polarisation

SFS Shape from Shading

SVD Singular Value Decomposition

Conventions

$*^{\intercal}$	Transpose operator
$\overline{*}$	Normalisation operator: $\frac{*}{\|\|*\|\|}$
$\boldsymbol{X}$	Vector $\in \mathbb{R}^k,\ k \geq 3$ (Bold, Uppercase)
$\boldsymbol{x}$	Vector $\in \mathbb{R}^2$ (Bold, Lowercase)
$\boldsymbol{X}^k$	k^{th} component of $\boldsymbol{X}$ (Bold, Uppercase and Exponent)
X	Matrix (Not-Bold, Uppercase)

Other Symbols

ϕ	Light source brightness magnitude
$\boldsymbol{H}$	Half vector
$\boldsymbol{L}$	Lighting vector
$\boldsymbol{N}$	Normal vector
$\boldsymbol{V}$	Viewing vector
$\boldsymbol{W}$	Weighted half vector
f	Focal length
u, v	Image plane coordinates
x, y, z	3D world coordinates

Chapter 1

Introduction

1.1 Objective

One of the main research aims of computer vision is 3D reconstruction, which is the process of digitising the shape and appearance of real world objects through algorithms using various sensors such as RGB cameras, active illumination systems, laser scanners etc. There are numerous approaches for the attainment of 3D reconstruction which have been the focus of study for decades. This dissertation will mostly centre on developing a 3D scanning technology aiming at:

- **Practical acquisition**: it is desirable to have a quick and practical data acquisition process without the need for expensive equipment (e.g laser scanners).
- **Dense Reconstruction**: it is also requisite that a surface representation as dense as possible is obtained so as to be able to recover surface details of sub-millimetre precision.
- **General surfaces:** it is furthermore imperative to be able to reconstruct surfaces of heterogeneous materials (including shiny ones) without the need for the existence of surface texture.

1.2 Motivation

Traditionally, 3D reconstruction has been used in a number of applications. Some examples include digitisation of several real world objects, such as archaeological artefacts. These 3D models can then be used in a variety of tasks such as virtual exhibitions or other entertainment applications, including but not limited to computer graphics for movies and video games.

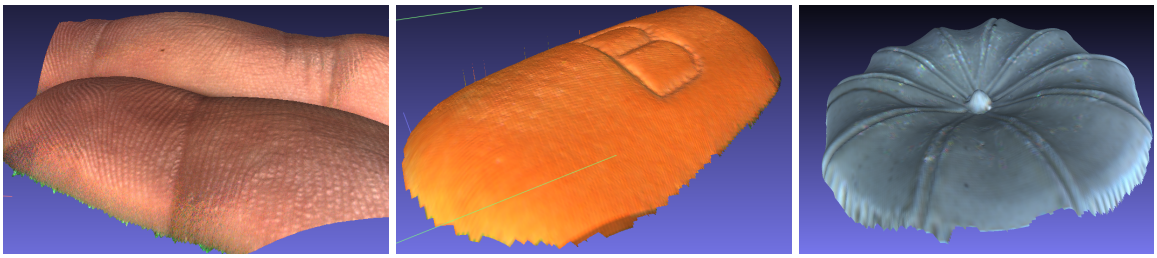


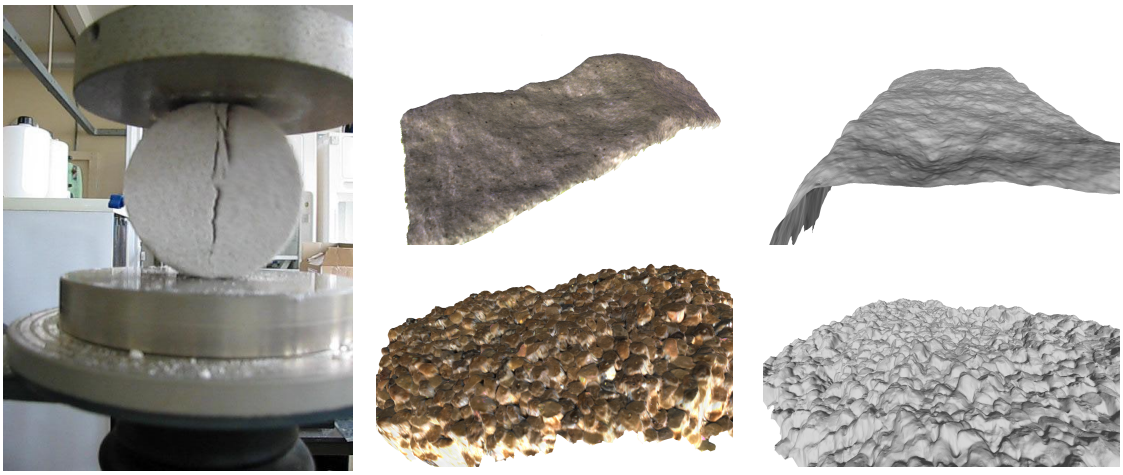
Figure 1.1 Example applications of the developed photometric-based 3D reconstruction system. (*left*) to (*right*) shows fingerprint scanning, inspection (aspirin pill) and heritage digitisation (tea caddy from the Fitzwilliam museum in Cambridge).

More recently, there has been an increasing need for augmented reality (AR) applications where a 3D model of a scene is used in conjunction with additional virtual objects superimposed. A closely related task is the navigation of autonomous robots, such as self-driving cars, where the 3D scan of the surrounding environment is required in order to safely interact with it.

In addition, with particular importance for this thesis are the tasks of inspection, for industrial as well as biometric applications. One example is a contact-less fingerprint scanner, utilising an accurate 3D reconstruction of a fingerprint simply from images, without the need for a specific touch sensor. Some examples of these applications are shown in Figure 1.1.

Another application of this work is the examination of the fracture of bio-cemented sandstone specimens. More specifically, artificial rocks generated through microbially induced calcite precipitation are of great importance for the geotechnical engineering community (e.g. in modelling the fracture of concrete, Kumar et al. [92]). These rocks are subjected to compression tests until fracture, as shown in Figure 1.2.

The usefulness of the 3D reconstruction system developed lies in that it offers a method for the quick examination of the surfaces through a very precise reconstruction so that the propagation of the crack in the material can be studied. In addition, the fast acquisition process (a few seconds for image capture and a few dozens of seconds for reconstruction) allows for the examination of a large number of samples, allowing for better experimental validation. For the work of Konstantinou et al. [90], 28 samples were scanned using the setup shown in Figure 1.5, with the whole acquisition process taking around 30 minutes using the GPU reconstruction algorithm discussed in Section 7.2.1.



(a) Compression test experimental setup.

(b) Surface reconstructions of the fragments.

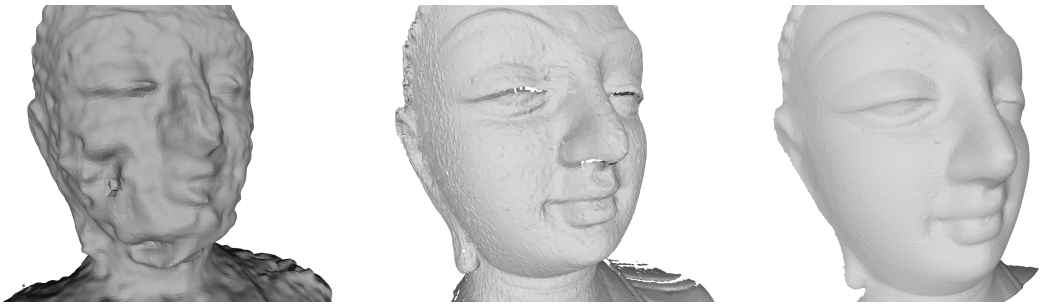
Figure 1.2 Investigation of the fracture of bio-cemented rocks. (b) shows a close-up textured and untextured of 3D reconstruction of the results of the fracture test performed with the setup of (a). This is an actual application of the contribution of this thesis appearing in the article (Konstantinou et al. [90]). For a reference of the scale of these samples, the diameter of the samples is a few centimetres, and the grains of the samples have average sizes of 0.18 and 1.8 mm for the fine grained (b-top) and coarse grained (b-bottom) samples respectively.

1.3 Background

The 3D reconstruction problem has been studied for decades and there are numerous approaches to it. A brief overview of different methods is presented below, and a comparison is shown in Figure 1.3.

As of 1981, one of the most remarkable pioneering 3D reconstruction works has been that of the eight point algorithm by Longuet-Higgins [107]. The eight point algorithm uses point correspondences in two images from two different points of view and allows simultaneous recovery of the camera motion between the two views as well as the corresponding 3D position of these points. This is the basis of a class of later algorithms called Structure-From-Motion (*SFM*). A similar setting, usually relevant to the robotics community is the Simultaneous Localization And Mapping (*SLAM*) problem, where the aim is usually to gather enough information about the surrounding 3D space so as to navigate an autonomous agent. Important work on SLAM includes Leonard and Durrant-Whyte [99], Klein and Murray [89], Engel et al. [44].

Given multiple images from different known camera positions, the 3D reconstruction problem is usually referred to as Multi-View Stereo (*MVS*). The special case of two images only is usually referred to as Binocular Stereo or just Stereo. In MVS, the objective is usually



(a) MVS. (b) Laser. (c) PS.

Figure 1.3 Comparison of 3D reconstruction methods. (a) was obtained using the VisualSFM pipeline ([193, 49, 84]) and it is of very poor quality due to the lack of surface texture. (b) was obtained using the laser scanner of NextEngine [131]. This is a good reconstruction although the surface is a bit noisy and of lower resolution than (c) which was generated using the proposed algorithm from Chapter 5.

to obtain an as dense as possible representation of the scene. MVS algorithms have had great success over the years (classic review papers Scharstein and Szeliski [160], Seitz et al. [162]) however their performance is usually dependent on the existence of point features such as surface texture, which can be matched across different views. Thus they are particularly suitable for reconstructing building-sized scenes with hundreds of images (e.g "Building Rome in a Day" project by Agarwal et al. [4] featuring large scale SFM and MVS) as opposed to a few images of a "small" (relative to the camera trajectory) object.

Another class of 3D reconstruction methods rely on structured illumination where a known pattern is projected on the scene; the deformation of the pattern, as captured by the camera, reveals the depth of the different objects of the scene. This is conceptually similar to the feature matching procedure of stereo algorithms, with the structured illumination providing most of the necessary features. The most characteristic example of such methods is the Kinect Fusion work by Newcombe et al. [130]. More recent extensions, also applicable for fusing data from generic active depth sensors (also time of flight) include Kerl et al. [87], Endres et al. [43] and Whelan et al. [189].

Finally, for this thesis, the most relevant family of 3D reconstruction algorithms rely on shading[1] to infer 3D structure. The foundation of this field is Horn [67], where the idea of Shape-From-Shading (*SFS*) was proposed and more relevantly Woodham [192], where the **Photometric Stereo** (*PS*) method was introduced. PS assumes multiple images from the same viewpoint along with varied illumination and calculates the surface orientation (normal) at each pixel through solving an inverse rendering problem. Then the surface normal map is

[1]The meaning of shading here is 'gradations of reflected light intensity' and should not be confused with the existence of shadows.

numerically integrated in order to produce a 3D surface. An example of PS input data and output is shown in Figure 1.4.

1.4 Photometric Stereo

(a) Input Images. (b) Normal map.

Figure 1.4 The Photometric Stereo method: (a) images from the same point of view but with varied illumination (here 2/8 are shown) are used to compute a normal map (b) and finally a 3D reconstruction (1.3c).

The operating principle of the Photometric Stereo method can be understood from the name itself. *Photometric* is a Greek word that means measuring the light. Indeed, it refers to the fact that the exact amount of reflected light (in practice measured as a pixel value) contains information about the surface. Although light reflection models can be very complicated as analysed in Section 2.1, in general, a high pixel value means that the surface normal is oriented to the incoming light direction whereas a low pixel value indicates that the surface normal is perpendicular to the incoming light. The term *Stereo* refers to the fact that the method critically depends on the change of illumination and so Stereo draws similarity with standard stereo techniques, where the change in the image information comes through motion. The need for more than one photometric image comes from the inherent ambiguity between shading and surface albedo. Albedo is an intrinsic surface property determining the total amount of light it reflects and is closely related to its perceived colour. In fact, there is a class of single image shape-from-shading methods which began with Horn [67] although these methods rely on a number of assumptions to overcome this ambiguity.

1.4.1 Motivation for Photometric Stereo

The Photometric Stereo technique is highly suitable for the acquisition of very detailed reconstructions. This is because it is a *direct method* that extracts information from every

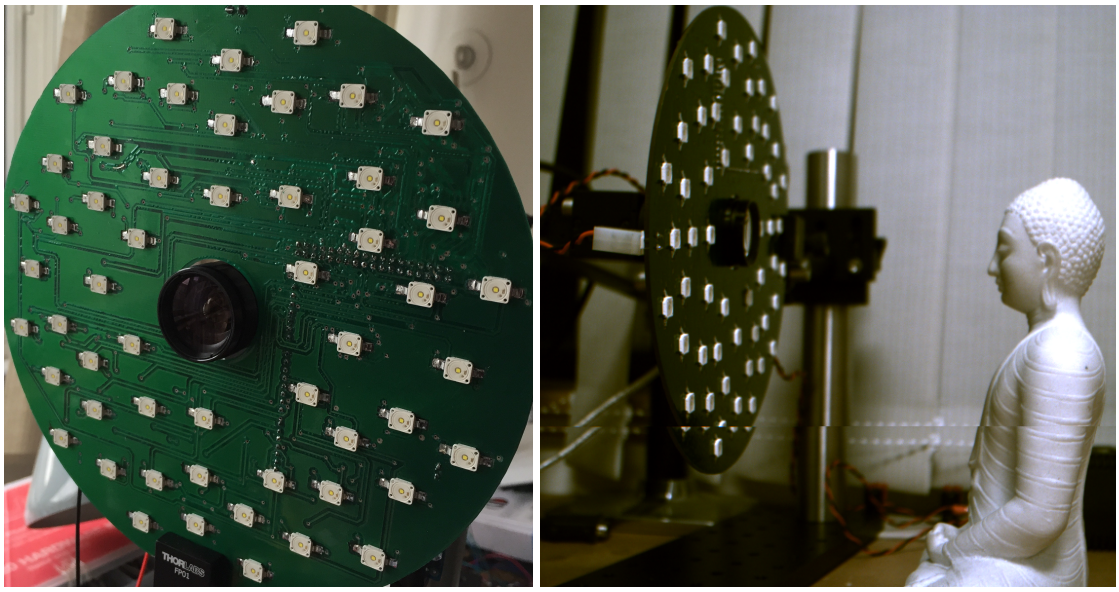


Figure 1.5 The hardware used for capturing data for the photometric-based reconstruction system developed. It consists of a simple circuit with inexpensive LEDs and a 3MPixel camera (FL3-U3-20E4C). This allows images to be captured with varied illumination from the same point of view and then use Photometric Stereo algorithms to obtain an object's 3D reconstruction. The setup contains 52 LEDs arranged in 6 co-centric circles in a 9cm radius disk. These circles have radii 3.5 to 8.5cm and they contain 6, 6, 6, 12, 12, 10 LEDs respectively. The LEDs in each circle are arranged uniformly with the aim to maximise the distance between nearby LEDs in successive circles. This arrangement of LEDs was designed in order to be able to generate a large number of lighting conditions, however, in practice, for most objects 6-9 images are enough for an accurate reconstruction.

image pixel without the need for prior extraction of key-points or other features. As a result, a large amount of information is readily available and thus 3D reconstruction meshes that have millions of vertices can be generated by using megapixel images. To add to this point, PS does not rely on texture or other specific surface characteristics; any non-black surface that reflects light is reconstructable, at least in theory. This makes PS applicable for a large variety of objects that are otherwise really hard to reconstruct with stereo matching algorithms reliant on texture and other distinct features. Moreover, it is noted that as the photometric information is encoded in the surface normals, a high level of detail is automatically recovered.

With regard to the aspect of implementation, PS allows for very fast acquisition as it does not involve any moving parts. Standard data acquisition involves 6-9 images and takes a few seconds (see Figure 1.5). This is orders of magnitude quicker than laser-scanning (e.g NextEngine [131]) which needs a few minutes to scan a sample, as it requires the laser to slowly move across the object's surface. In addition, despite the need for specific hardware, its material cost is minimal as LEDs cost a few pounds each and are controlled using a simple

Arduino controller. This is again contrasted with commercial laser scanners that cost at least a few thousand pounds.

Finally, the lack of moving parts property of PS acquisition makes it suitable for applications where moving is difficult such as endoscopy (Wu et al. [196]) or monitoring the development of glaucoma (Lee and Brady [98]).

1.4.2 Literature limitations

Woodham's classic PS method proved to be quite impractical as it relies on a set of restrictive assumptions about the illumination (parallel uniform), orthographic viewing geometry and diffuse surface reflection. All of these assumptions are invalidated when using the practical capture setup of Figure 1.5. Over the years, the extensive research on the matter has managed to overcome some of these limitations. This thesis goes a few steps further in Photometric Stereo modelling, allowing for a more practical system to be made.

The aim of this thesis is to extend the Photometric Stereo technique in order to make it more applicable and practical. It is imperative to be able to deal with objects close to the camera, as seen in Figure 1.5, for the following two reasons. Firstly, the closer the camera gets to the object, the more surface details are visible and hence can be reconstructed. In addition, placing the target object close to the light sources allows them to outshine any ambient lighting and hence make acquisition not require a dedicated dark room. In general, the closer to the light sources the object is, the better signal-to-noise ratio can be obtained. As a result, in order to perform nearby acquisition, perspective projection as well as point light sources need to be modelled.

Finally, it is required that objects of reflectance as general as possible and not just diffuse be dealt with. This includes applicability to heterogeneous scenes with multiple objects and complicated geometry.

From a methodology approach, the aim is the mathematical manipulation of the equations governing the image formation process (irradiance equation, see Section 2.1) in order to get them into more computationally tractable forms. This is necessitated as light reflection models can be very non-linear and thus make the inverse rendering problem unsolvable. In addition, as illumination crucially depends on the surface normal, it is highly desirable to express the Photometric Stereo problem in a differential form, with the differentiated variable being the depth away from the image plane. This differential parameterisation allows for the use of mathematical tools from the partial differential equation (*PDE*) literature as well as powerful variational solvers, allowing for high levels of robustness in real world imperfections such as camera thermal noise.

Finally, it has to be noted that it is a crucial aim of the thesis to develop a general mathematical framework that can reliably model as many situations as possible. This means that it is desirable to use minimal prior knowledge about the target scenes. This limitation also restricts the applicability of machine learning approaches (either deep or classical) as it is definitely preferable to avoid using training data as this would bias the applicability towards objects similar to the training set.

1.4.3 Challenges in Photometric Stereo modelling

Most of the challenges in computer vision, and more specifically in the 3D reconstruction sub-field, arise from the fact that it mostly deals with *inverse* problems. The physics of how light reflects off object's surfaces, is then scattered by the atmosphere, refracted through camera lenses (or human eyes), and finally projected onto a flat (or curved) image plane is well understood and computer graphics are very successful at simulating these processes. However, the inverse problem has proved to be much harder. Going from 2D images to 3D representations of the world is a non-trivial, and an ill-constrained problem, as the dimensionality of the output (objects in 3D space with additional colour and reflectance properties) is far greater than the dimensionality of the input (a set of 2D images). Therefore, the modelling process is a trade-off between capturing enough real world features and at the same time having solvable equations, with numerically tractable solutions.

1.5 Contributions

The contribution of the thesis is threefold. First of all, the Photometric Stereo model is extended in order to be able to deal with arbitrary ambient lighting. This is a step towards acquisition in a not fully controlled lab setting. Secondly, the need for a priori knowledge of the light source brightness and attenuation characteristics is relaxed as an alternating optimisation procedure is proposed which is able to estimate these parameters. This extension allows for quick acquisition with cheap LEDs that exhibit unpredictable illumination characteristics (flickering, etc). Finally, a volumetric parameterisation is proposed which allows one to tackle the multi-view Photometric Stereo problem in a very similar manner, in a simple unified differential model. This final extension allows for complete object reconstruction merging information from multiple images taken from multiple viewpoints and variable illumination.

The relevant articles are:

- [105] Logothetis, F., Mecca, R., Quéau, Y., and Cipolla, R. (2016). Near-field photometric stereo in ambient light. In *British Machine Vision Conference (BMVC)*
- [103] Logothetis, F., Mecca, R., and Cipolla, R. (2017). Semi-calibrated near field photometric stereo. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 3, page 8
- [104] Logothetis, F., Mecca, R., and Cipolla, R. (2018a). A differential volumetric approach to multi-view photometric stereo. *arXiv*, 1811.01984

In the aforementioned works, I contributed to the theoretical framework and also made almost all of the coding implementation. In addition, I have also contributed to the following publications as part of my work on Photometric Stereo modelling:

- [118] Mecca, R., Quéau, Y., Logothetis, F., and Cipolla, R. (2016). A single lobe photometric stereo approach for heterogeneous material. *SIAM Journal on Imaging Sciences*, 9(4):1858–1888
- [116] Mecca, R., Logothetis, F., and Cipolla, R. (2017). A differential approach to shape from polarization. In *British Machine Vision Conference (BMVC)*
- [106] Logothetis, F., Mecca, R., Sgallari, F., and Cipolla, R. (2018b). A differential approach to shape from polarisation: a level-set characterisation. *International Journal of Computer Vision (IJCV)*, pages 1–14

For [118], I ran the experiments presented in Figures 3.2 and 3.4. For [116], I contributed at the implementation part and finally in [106], I also contributed to the theoretical framework extension compared to [116]. Finally, the theory developed in Chapter 5 along with the experimental setup were used in order to provide scanning for the following Geotechnical Engineering paper:

- [90] Konstantinou, C., Biscontin, G., and Logothetis, F. (2019). Tensile strength of bio-cemented sands. Under Review for Journal of rock mechanics and rock engineering

1.6 Outline

The rest of the thesis is structured as follows.

Chapter 2 gives a more in depth discussion of the classical Photometric Stereo problem and the pioneering solutions to it. A detailed review of the classical literature about light reflection modelling and numerical integration methods is also included.

Chapter 3 describes the modern theoretical framework which is the basis for the Photometric Stereo modelling of the thesis. This includes PDE manipulation using the image ratio method and modelling from the applied mathematics community, as well as a variational optimisation framework.

Chapter 4 describes the extension of the Photometric Stereo method to handle arbitrary ambient light. This additional unknown complicates the irradiance equation to the point that the ratio method is made inapplicable. The solution is to use a binomial series expansion to linearise the irradiance equation and then consider a ratio of differences in order to derive a simplified formulation.

Chapter 5 introduces the semi-calibrated Photometric Stereo problem, where the light source positions are known though the brightness is not. This is a very practical issue arising from high speed acquisitions where LEDs tend to flicker. A solution to this is presented by the means of an alternating optimisation framework.

Chapter 6 extends the PS framework into a multi-view setting. This allows for complete 3D reconstructions for scenes with discrete objects in contrast to the continuous depth maps that have been generated for the rest of the thesis. Indeed, this extension requires a different parameterisation based on the signed distance field, considered along with its relationship to the surface normal in order to propose a novel differential parameterisation.

Chapter 7 concludes the thesis, also showing how PS modelling can be merged with the shape-from-polarisation method. This is a very interesting application showing the general applicability of the theoretical framework developed in this thesis.

Chapter 2

Background

This chapter provides the basic background for the thesis. Firstly, some light reflection properties are presented along with how they have been exploited in the computer graphics literature. Then, the classic Photometric Stereo and the shape from shading algorithms are analysed as these are the foundations of the field. Finally, the chapter concludes with a presentation of some recent advances in deep learning for Photometric Stereo problems.

2.1 Light reflection basics

This section presents some notions of light reflection in physics which are at the core of the image formation process. For a more detailed explanation refer to Nicodemus [133], Szeliski [173].

In general, the scattering of light off a surface depends on its *Bidirectional Reflectance Distribution Function (BRDF)*. The BRDF is defined as the ratio of the incoming *radiance* to the reflected one. The radiometric quantity *radiance* measures the amount of light power per projected area and per steradian, reflected, transmitted or received by a given surface; it corresponds roughly to the amount of light power transmitted per ray. In general, a BRDF is a function of four variables parameterising two unit length vectors $\overline{\boldsymbol{L}}$ and $\overline{\boldsymbol{V}}$. The operator $\bar{\cdot}$ is used to define normalisation of a vector as:

$$\overline{\boldsymbol{X}} = \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \tag{2.1}$$

Vectors $\boldsymbol{V}$ and $\boldsymbol{L} \in \mathbb{R}^3$ denote the viewing and lighting vectors as defined in the caption of Figure 2.1.

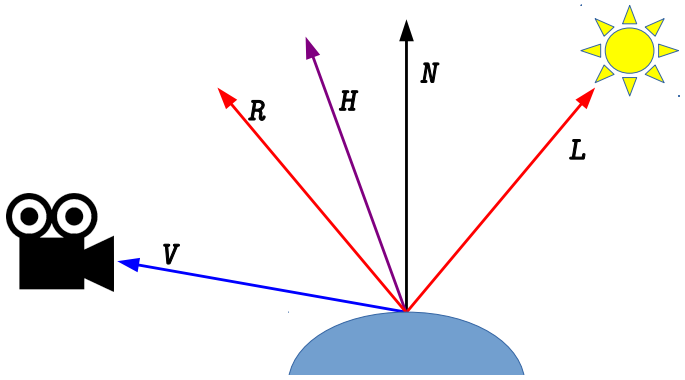


Figure 2.1 Most relevant vectors for modelling light reflection and BRDFs. These are the surface normal $\boldsymbol{N}$, lighting vector $\boldsymbol{L}$ pointing towards the light source, the viewing vector $\boldsymbol{V}$ pointing towards the camera, $\boldsymbol{H}$ the half-vector between $\boldsymbol{V}$ and $\boldsymbol{L}$ ($\boldsymbol{H} = \overline{\boldsymbol{V} + \boldsymbol{L}}$) and finally $\boldsymbol{R}$, the direction of mirror-like reflection ($\boldsymbol{N} = \overline{\boldsymbol{L} + \boldsymbol{R}}$).

In principle, the BRDF of a real surface satisfies the following constraints:

1. **Non-negativity:** For any direction, the amount of reflected light is non-negative.
2. **Conservation of energy:** The total amount of reflected light equals the incoming light.
3. **Helmholtz reciprocity:** The BRDF is symmetric between swapping incoming and outgoing directions.

Although the BRDF is defined through a ratio of powers at the surface interphase, it is very common to assume that light propagates unattenuated through the air, so that the incoming radiance to the surface equals the radiance emitted by the light source and the reflected radiance equals the radiance received by the image sensor. It must be noted that these assumptions are severely violated in attenuating media (such as murky waters [179, 181, 126]) but this is beyond the scope of this thesis. Thus, assuming that the camera has a linear response (so that the pixel values are proportional to the incoming radiance), the BRDF leads to an *image irradiance equation*, where each pixel value is expressed with the BRDF for the camera view vector $\boldsymbol{V}$ and the lighting vector $\boldsymbol{L}$ as:

$$i(u,v) \propto \mathrm{BRDF}\Big(\boldsymbol{V}(u,v), \boldsymbol{L}(u,v)\Big) \tag{2.2}$$

where vectors $\boldsymbol{V}(u,v)$ and $\boldsymbol{L}(u,v)$ are now expressed in image-centred coordinates. The proportionality in Equation 2.2 is used to signify that the exact image value depends on a number of other constant parameters such as the camera's analogue gain and the intrinsic brightness of the light source.

The straightforward way of measuring a real BRDF involves the positioning of a light source and a detector at various directions from an object of simple geometry, usually a flat sample of the material to be measured, Marschner et al. [113]. Alternatively, a half-silvered mirror and a digital camera can capture many BRDF samples of a planar target at once, Ward [186]. However, dealing with general BRDFs is very hard in practice, especially from a PS point of view, where the inverse problem is required (Section 1.4.3). Thus, some analytical models approximating BRDFs are preferred and this is explored in the following sections.

Before proceeding, it is noted that another common method of compressing the information of the BRDF is to project it onto spherical harmonics, e.g. Westin et al. [188] and Ghosh et al. [52]. These are the spherical analogues of sines and cosines, in that they form a smooth orthonormal basis for functions on the sphere. Simple BRDFs have few non-negligible spherical harmonic coefficients, thus truncating the series to a few terms is a good approximation.

A generalisation of the BRDF is the *Bispectral Bidirectional Reflectance and Reradiation Distribution Function (BBRRDF)*, Hullin et al. [69] applicable for fluorescent materials, where a portion of the incident light is absorbed and re-emitted at a higher wavelength. Treibitz et al. [178] suggest and experimentally verify that fluorescence emission is well approximated by an ideal Lambertian model (see Section 2.1.1) and so the classic PS method (see Section 2.2) is mostly applicable. A similar approach is also followed by Sato et al. [159].

2.1.1 Lambertian reflection

The simplest BRDF is the *Lambertian* dating back to the 18th century and published in Lambert [96]. This function simply assumes that once light reflects off an object it spreads uniformly in all viewing directions. In addition, the amount of reflected light is proportional to the cosine between the surface normal and the lighting vector. This leads to the following irradiance equation for the diffuse component of the reflected light i_d:

$$i_d(u,v) \propto \max\left(\overline{\boldsymbol{N}}(u,v) \cdot \overline{\boldsymbol{L}}(u,v), 0\right) \tag{2.3}$$

where the max operator is used to signify that those points where the surface normal is faces away from the surface, receive no light. This case corresponds to an attached shadow and will be further examined in Section 2.2.1. For now, as it is usually the case in the literature, the max operation will be ignored although it always applies implicitly, as BRDFs are always non-negative.

Surprisingly, this very simple Lambertian model is quite accurate for a range of materials experiencing diffuse reflections and it forms the basis of most of the PS literature (e.g [192, 66, 138, 72, 144].

2.1.2 Analytic reflection models

(a) Lamberian [96] (b) Blinn–Phong [17] (c) Cook and Torrance [32]

Figure 2.2 Comparison of different light reflection models for rendering synthetic data.

The Lambertian BRDF model presented in Section 2.1.1 completely ignores any kind of *specular* reflection and thus is limited, at least from a rendering perspective. Specular reflection is a mirror-like reflection where a significant portion of the incoming light is reflected around a specific direction only, usually around the $\boldsymbol{R}$ vector (see Figure 2.1) depending on the material. On the other hand, look-up tables are usually too computationally expensive to use in real-time rendering tasks. The middle ground is to approximate the BRDF with some analytical model. Many attempts in direct analytical models have been proposed, as shown in the surveys by Ngan et al. [132] and Kurt and Edwards [93].

One of the earliest models, still considered useful today, was presented in Phong [146]. It directly exploits the empirical observation that for various materials most of the specular reflection is around $\boldsymbol{R}$ (see Figure 2.1) and proposes a power law fall off as $\boldsymbol{V}$ moves away from $\boldsymbol{R}$. Another popular alternative is the Blinn–Phong model [17] which sets the specular lobe around the half vector $\boldsymbol{H}$. Phong and Blinn–Phong model formulas are shown in Equations 2.4 and 2.5 respectively:

$$i_p \propto (\overline{\boldsymbol{V}} \cdot \overline{\boldsymbol{R}})^{c_p} \tag{2.4}$$

$$i_b \propto (\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{H}})^{c_b} \tag{2.5}$$

where c_p and c_b are material constants determining how strong the reflection is; the higher the value of the exponent, the stronger the specular reflection is, which corresponds to smaller, brighter highlights. The limiting case of infinity exponents corresponds to a Dirac delta

function and this has been considered as reflectance model in [128] (more about the usage of reflection models in PS in Section 3.4.1).

More physically-based BRDFs include the Torrance and Sparrow [176], the Cook and Torrance [32], the Ward [186] and Lafortune et al. [94]. The Cook & Torrance (CT) model [32] equation is presented here as it is used later:

$$i_{ct} \propto \frac{fdg}{\pi(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{L}})(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{V}})} \quad \text{where:} \tag{2.6}$$

$$f = f_0 + (1 - f_0)(1 - (\overline{\boldsymbol{V}} \cdot \overline{\boldsymbol{H}}))^5 \tag{2.7}$$

$$f_0 = (\frac{n_1 - n_2}{n_1 + n_2})^2 \tag{2.8}$$

$$d = \frac{1}{\pi r^2 (\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{H}})^4} e^{\left(\frac{(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{H}})^2 - 1}{r^2 (\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{H}})^2}\right)} \tag{2.9}$$

$$g = \min\left(1, \frac{2(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{H}})(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{V}})}{(\overline{\boldsymbol{V}} \cdot \overline{\boldsymbol{H}})}, \frac{2(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{H}})(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{L}})}{(\overline{\boldsymbol{L}} \cdot \overline{\boldsymbol{H}})}\right). \tag{2.10}$$

More specifically, the CT model is based on the idea that the surface contains microfacets (originally introduced in [176]), acting like perfect mirrors. Then the f term (not to be confused with the focal length) models light refraction (n_1 and n_2 are the indices or refraction of the material and the medium which is usually the air; the formula above is a simplification proposed by Schlick [161]), d models the microfacet distribution (i.e. the number of microfacets aligned with the half vector $\boldsymbol{H}$) and g models the mutual shadowing between microfacets. An example rendering that compares purely diffuse to purely Blinn–Phong and to purely Cook and Torrence is shown in Figure 2.2.

Finally, note that a popular graphics strategy (e.g. [95]) is to mix the BRDFs described above with a Lambertian component, so for example $i_{all} = \rho_d i_d + \rho_{ct} i_{ct}$, with $\rho_d + \rho_{ct} = 1$

2.2 Classical Photometric Stereo

In this section, a more detailed explanation of the Photometric Stereo (PS) method introduced by Woodham in [192] is given. PS assumes images from the same viewpoint as well as varied but known illumination. This classic method relies on the following set of assumptions

1. Lambertian Reflection.
2. Orthographic projection.
3. Uniform Known Parallel illumination.

4. No shadows or self-reflections.

5. Linear sensor (camera) response.

Under the aforementioned set of assumptions the classic PS irradiance equation for pixel $i(u, v)$ for light source j is:

$$i_j(u, v) = \rho(u, v)(\overline{\boldsymbol{N}}(u, v) \cdot \boldsymbol{L}_j) = \boldsymbol{L}_j^\intercal \rho(u, v)\overline{\boldsymbol{N}}(u, v). \tag{2.11}$$

Now, the proportionality of Equation 2.3 is replaced with an equality with the following two changes: firstly, the pixel-wise unknown $\rho(u, v)$ is included. This is called *albedo* which is a scalar associated with the reflectance of the surface at that particular point. Albedo is closely related to the intrinsic color of the surface. Secondly, the normalisation of the lighting vector $\boldsymbol{L}$ is removed to allow for the possibility of light sources with difference intrinsic brightness (encoded on $\|\boldsymbol{L}\|$) [1].

Note that the orthographic projection assumption guarantees that points on the image plane (u, v) directly map to 3D points as $x(u, v) = u$ and $y(u, v) = v$ thus it makes sense to express the irradiance of a 3D point through quantities calculated at the image plane. Assumption 4 guarantees that the pixel values only correspond to the reflected light (through Lambertian reflection).

The objective of the classic PS problem is to recover $\rho(u, v)$ and $\boldsymbol{N}(u, v)$. The baseline solution involves the use of the auxiliary variable $\boldsymbol{M}(u, v) = \rho(u, v)\overline{\boldsymbol{N}}(u, v)$ and then stacks different irradiance Equations 2.11 that come out of images taken under different illumination $\boldsymbol{L}_j$ for $j = 1, \dots n$. This leads to the following linear system:

$$\begin{bmatrix} \boldsymbol{L}_1^\intercal \\ \vdots \\ \boldsymbol{L}_n^\intercal \end{bmatrix} \boldsymbol{M}(u, v) = \begin{bmatrix} i_1(x, y) \\ \vdots \\ i_n(u, v) \end{bmatrix} \quad \text{that is} \tag{2.12}$$

$$L_s \boldsymbol{M}(u, v) = \boldsymbol{I}_s(u, v). \tag{2.13}$$

The system is solvable assuming that the rank of L_s is 3. This requires $n \geq 3$ as well as non-collinear light sources. The least squares solution is $\boldsymbol{M}(u, v) = (L_s^\intercal L_s)^{-1} L_s^\intercal \boldsymbol{I}_s(x, y)$ which in turn is used to calculate $\rho(u, v) = \|\boldsymbol{M}(x, y)\|$ and $\boldsymbol{N}(u, v) = \frac{\boldsymbol{M}(u,v)}{\|\boldsymbol{M}(u,v)\|}$.

[1] [192] only considers normalised light vectors but the solution strategy is identical for $\boldsymbol{L}$ of any magnitude. In fact, assuming normalised lighting vectors simplifies the uncalibrated PS problem discussed in section 5.2.1).

2.2.1 Robustness to outliers

Since the work of [192], the field has progressed. There are two different categories of progress. One is to extend the model in order to parameterise more physical phenomena. This will be analysed in depth in Chapter 3. The second category of methods focuses on solving PS under the setting where the classic assumptions are mostly true, aiming to maximise robustness to outliers.

Outliers are data points which, in a statistical sense, are significantly different from the model and hence require special treatment. The baseline method of solving Equation 2.13 in the least-squares sense is only justified in the presence of Gaussian noise only, which is not a realistic assumption, and it is sensitive to outliers. It is usually assumed that outliers are sparse: this means that a relatively small percentage of data points are outliers and that the rest follow the model under some sort of noise assumption. Hence, it is common to assume that specular reflections, which are significant for points where the surface normal is aligned with $\boldsymbol{R}$ or $\boldsymbol{H}$, (see Section 2.1) are sparse outliers. In addition, shadows caused

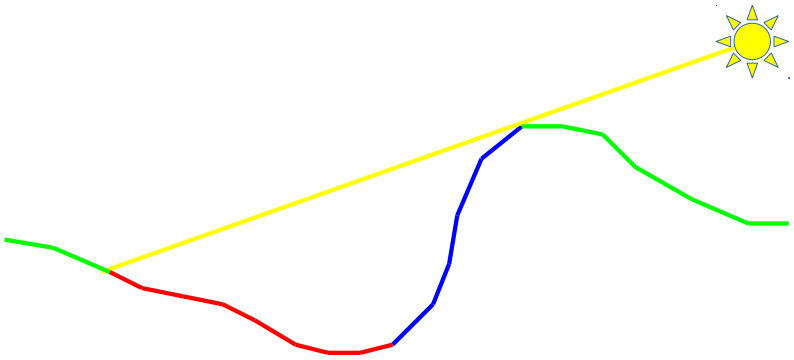


Figure 2.3 Illustrations of attached and cast shadows which are (along with specular highlights) the main outliers in PS. Attached shadows (blue parts of the surface) occur when the surface normal is directed away from the light source (i.e. $\boldsymbol{N} \cdot \boldsymbol{L} < 0$) and thus the surface shadows itself locally. On the contrary, cast shadows (red parts of the surface) occur when one part of the surface occludes another part of it. This is a global illumination effect and it is much harder to detect in practice.

by self-occlusions of the surface (see Figure 2.3) are also one of the most common sparse outliers. Note that near lighting, perspective viewing geometry and ambient light cannot be assumed as outliers as these effects dramatically alter the irradiance equation for every pixel in every image. Hence, these effects cannot be managed with robust estimations and have to be specifically modelled. This will be discussed in Sections 3.3 and 4.2.

To specifically model these sparse corruptions, Wu et al. [199] proposed a re-written version of Equation 2.13 as:

$$I = L_s M + E \tag{2.14}$$

where I and M are the matrices consisting of stacking $\boldsymbol{I}_s(u, v)$ and $\boldsymbol{M}(u, v)$ and E is the corruption matrix. Equation 2.14 is "hopelessly under-constrained"[2] unless we assume that E is very sparse and most of its elements are 0. The natural way to express this is to try to minimise its cardinality, i.e. $\|E\|_0$. However, such a minimisation is NP hard and so this has to be approximated in practice. A common approximation is to use the ℓ_1 norm i.e. $\|E\|_0 \approx \|E\|_1$ as used, for example, by Wu et al. [199] to solve the problem using low-rank factorisation. Similarly, Ikehata et al. [72] follow sparse Bayesian regression for the same constraint. Wang et al. [185] proposed dealing with the non-convex nature of the low-rank matrix factorisation problem through a proximal strategy. Finally, [59, 70] eliminated outlier points using RANSAC ([47]).

In addition, it is widely considered true that robustness is increased if the PS problem is posed in a differential form (with the depth as unknown) and resolved through a variational optimisation. This will be further examined in the next section.

Moreover, in order to further increase robustness, it is common to include a *robust kernel function* f. Thus a minimisation of an error function $\min_x e(x)$ is re-written as $\min_x f(e(x))$. This transformation aims to minimise the effect of outliers (which have high values for $e(x)$) which would otherwise dominate the error function. A few common kernel functions include:

- ℓ_2 norm: $f(x) = 0.5x^2$
- ℓ_1 norm: $f(x) = |x|$
- Huber: $f(x) = \begin{cases} 0.5x^2, & \text{if } |x| \leq \lambda \\ \lambda(|x| - 0.5\lambda) & \text{if } |x| > \lambda \end{cases}$, λ is a free parameter.
- LogCosh: $f(x) = \log(\cosh(x))$.
- Cauchy: $f(x) = \lambda^2 \log\left(1 + \frac{x^2}{\lambda^2}\right)$, λ is a free parameter.

These kernel functions have different advantages and disadvantages. In general, the set of desirable properties include differentiability, convexity, simple analytic inverse and

[2] The phrase "hopelessly under-constrained" will be used throughout the thesis to refer to a problem where the number of unknowns vastly outnumbers the number of equations/constraints and thus a solution requires additional considerations. This is in contrast to a "simple under-constrained" problem where a solution can be found up to some "simple" ambiguity. For example, a linear system with n unknowns and $n - 1$ equations has 1 parameter ambiguity whereas a system with $2n$ unknowns and n equations is "hopelessly under-constrained".

$\lim_{|x|\to\infty} \frac{\vartheta f(x)}{\vartheta x} = 0$ (so that the influence of outliers is minimised) and the above functions have different trade-offs in these objectives. An illustration of these different kernel functions is shown in Figure 2.4.

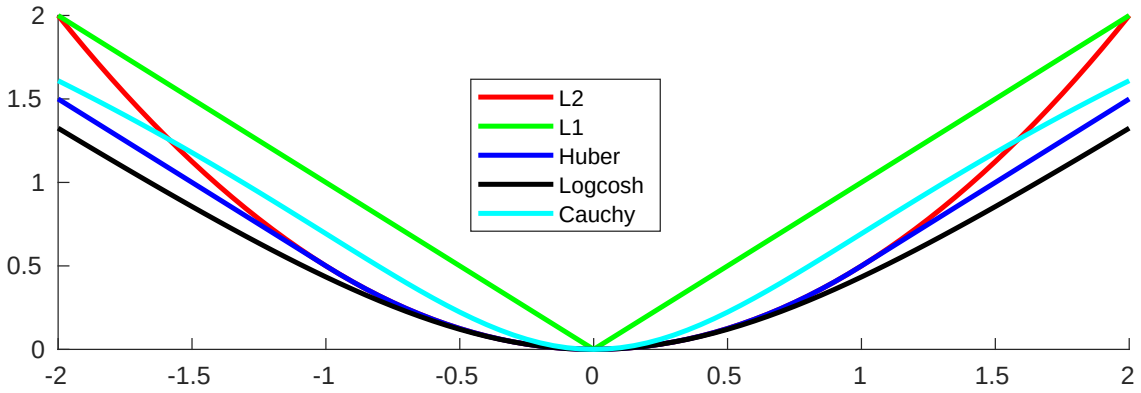


Figure 2.4 Illustration of the 5 different kernel functions with formulas given above ($\lambda = 1$ is used for the free parameter).

Robust variational optimisation for the normal integration problem include Quéau and Durou [150], Durou et al. [41] and finally Quéau et al. [154] who compared several robust loss functions for the direct depth estimation problem.

2.2.2 Integration of the normal field

Section 2.2 presented the classical PS method which is able to generate a map of the surface normal orientation $\boldsymbol{N}(u, v)$ as well as the surface albedo $\rho(u, v)$. These surface properties are very useful for a number of tasks (e.g. in graphics). However, in general, it is more practical to have a 3D reconstruction. There are several ways of parameterising a 3D object. The most straightforward way for monocular techniques like PS is a depth map, i.e $z(u, v)$. Thus the 3D object becomes the set of 3D points $[x(u, v), y(u, v), z(u, v)]$ with $x(u, v) = u$ and $y(u, v) = v$ assuming orthographic camera viewing.

Obtaining a depth map out of a normal map is a *numerical integration* problem, as the surface normal is fully specified from the depth partial derivatives (assuming that the depth map $z(u, v)$ is differentiable). For orthographic projection, the depth/normal relation is:

$$\boldsymbol{N}(u, v) \propto \left[\frac{\partial z(u, v)}{\partial u}, \frac{\partial z(u, v)}{\partial v}, -1\right]^{\mathsf{T}}. \tag{2.15}$$

Although not always true for real surfaces (occlusion boundaries cause depth discontinuities, sharp edges correspond to non-differentiable parts of the surface) it is assumed that

the normal field follows the integrability constraint, that is $\frac{\partial z(u,v)}{\partial u \partial v} = \frac{\partial z(u,v)}{\partial v \partial u}$ (Horn and Brooks [66]). Substituting that into Equation 2.15 and using $\boldsymbol{N}(u,v) = [\boldsymbol{N}^1, \boldsymbol{N}^2, \boldsymbol{N}^3]^\intercal$ leads to:

$$\frac{\partial(\frac{\boldsymbol{N}^2}{\boldsymbol{N}^3})}{\partial u} = \frac{\partial(\frac{\boldsymbol{N}^1}{\boldsymbol{N}^3})}{\partial v}. \tag{2.16}$$

In practice, Equation 2.16 is never exactly satisfied for calculated normal maps, even for points far away from discontinuities; thus recovering z from $\boldsymbol{N}$ is non-trivial. A survey of classic and modern methods as well as extensions is found in Bähr et al. [8]. Although modern PS methods solve for z directly, skipping the numerical integration step, a quick overview of some baseline numerical integration methods is presented here.

First of all, the normal integration problem can be thought of as a subcategory of problems that can be expressed as solving the *Eikonal Equation*, namely:

$$\|\nabla t(u,v)\| = \frac{1}{f(u,v)} \tag{2.17}$$

where $t(u,v)$ is the unknown scalar field in question and $f(u,v)$ is known.

One of the fastest ways of solving Equation 2.17 is the fast-marching approach introduced in Sethian [164]. This relies on the fact that information flows outwards like a propagating waveform; thus the function t can be propagated from a single point to the whole domain in a wave-like function. Normal integration for Photometric Stereo was first expressed in eikonal equation form in Ho et al. [63]. They introduced the auxiliary variables $p = \frac{\partial z}{\partial u} = -\frac{n^1}{n^3}$ and $q = \frac{\partial z}{\partial v} = -\frac{n^2}{n^3}$ thus, $\|\nabla z\| = \sqrt{p^2 + q^2}$ is of the form of Equation 2.17 and can be solved quickly with fast-marching.

An alternative way of numerical intergeneration using the Fourier Transform was introduced by Frankot and Chellappa [48]. Using the p, q definitions from above and performing differentiation, the integration problem can be expressed in the Poisson equation form as:

$$\nabla^2 z(u,v) = \frac{\partial p}{\partial u} + \frac{\partial q}{\partial v}. \tag{2.18}$$

Then, performing Fourier transform FT: $f(u,v) \rightarrow \mathrm{FT}(f(\omega_u, \omega_v))$, in both sides of Equation 2.18 and using the property that $\mathrm{FT}\big(\frac{\partial f(u)}{\partial u}\big) = j\omega_u \mathrm{FT}(f(u))$, the problem is transformed as:

$$(\omega_u^2 + \omega_v^2)\mathrm{FT}(z(u,v)) = j\omega_u \mathrm{FT}(p(u,v)) + j\omega_v \mathrm{FT}(q(u,v)). \tag{2.19}$$

Dividing by $(\omega_u^2 + \omega_v^2)$ (assuming it as non-zero) and taking the inverse Fourier transform IFT in Equation 2.19 leads to the following analytic solution for $z(u, v)$:

$$z(u, v) = \text{IFT}\Big(\frac{j\omega_u \text{FT}(p(u, v)) + j\omega_v \text{FT}(q(u, v))}{(\omega_u^2 + \omega_v^2)}\Big). \tag{2.20}$$

We note that the division by $(\omega_u^2 + \omega_v^2)$ becomes invalid for the zeroth frequency, corresponding to the mean value of z. This is because every mean value of z is consistent with Equation 2.19 and this has to be determined with some other measurement. In addition, Equation 2.20 provides some justification for the observation that the depth maps generated by PS methods usually suffer from low frequency deformations while keeping high accuracy at high frequencies. This is because of the $\frac{j\omega_u}{\omega_u^2+\omega_v^2}$ and $\frac{j\omega_v}{\omega_u^2+\omega_v^2}$ terms that scale errors in lower frequencies more than in high frequencies. Even if some other integration strategy is used (e.g. the fast marching approach described above or even the variational solver described in Section 3.7.7), Equation 2.19 is still approximately true and hence the inaccuracy at low frequencies is an inherent limitation of most differential approaches.

Finally, we note that Equation 2.19 can be modified using the Discrete Cosine Transform (DCT) as proposed by Simchony et al. [168]. This avoids the implicit assumption of periodic boundary condition that FT introduces.

Equation 2.20 (or the equivalent DCT formulation) can be computed very quickly with the fast Fourier transform algorithm. However, its disadvantage is that it is only applicable to rectangular domains; applying it to a real image with an arbitrary segmentation mask will cause severe bias from zeros in out-of-domain regions.

2.3 Shape from shading

The idea of linking photometric cues to 3D shape is older than the Woodham's classic Photometric Stereo paper [192]. The original observation comes from the astronomy community with the works of Van Diggelen [182], Rindfleisch [156], who suggested using the shading in Lunar images in order to infer the local surface height. This approach was later formalised and generalised in Horn's PHD thesis [67] who also suggested the term "shape-from-shading"(SFS) and thus he is generally regarded as the founder of the field.

The classical SFS problem has a very similar formulation and set of assumptions with those of the classical PS problem (section 2.2). The most important difference is that the input is a single intensity image and thus the albedo has to be considered constant (or equivalently known). Otherwise, the problem is "hopelessly ill-constrained" and any possible surface can be a potential solution given a corresponding albedo map. However, even under the constant

albedo assumption, the SFS constraint is not enough to uniquely specify the surface normal, as the irradiance equation for SFS is

$$i(u, v) = \overline{\boldsymbol{N}}(u, v) \cdot \overline{\boldsymbol{L}}. \tag{2.21}$$

In Equation 2.21, $\boldsymbol{N}(u, v)$ is uniquely specified only in a finite set of points (called singular) where $i(u, v) = 1$ and thus $\boldsymbol{N}(u, v) = \overline{\boldsymbol{L}}$ and for the rest of the domain there is a one parameter ambiguity. To overcome this ambiguity, Horn [67] suggested expressing the SFS problem in a differential form. This involves the normal parametrisation described in Section 2.2.2 using the partial depth derivatives p, q.

$$i = \frac{pl^1 + ql^2 - l^3}{\sqrt{p^2 + q^2 + 1}} \iff i\sqrt{|\nabla z|^2 + 1} + [l^1, l^2] \cdot \nabla z - l^3 = 0 \tag{2.22}$$

where $[l^1, l^2, l^3]^\intercal$ is the light source vector.

A very common subcategory of Equation 2.22 is to assume a frontal light source at infinity $[0, 0, 1]$, thus the SFS problem reduces to $|\nabla z| = \sqrt{\frac{1}{i^2} - 1}$ which is the Eikonal equation form (Equation 2.17). [67]'s solution to the SFS problem included the characteristic strip expansion method: this involves re-writing the non-linear PDE as a system of five linear first order ODEs for the variables $u(t) = x(t)$, $v(t) = y(t)$ (orthographic projection), $z(t)$, $p(t)$ and $q(t)$, where t is the parameterisation of the characteristic lines. Thus the problem has a unique solution assuming that all of these five variables have continuous derivatives which translates to C2 continuity[3] for the original surface. In addition, the value of all variables is needed at the boundaries (Dirichlet boundary condition) which translates to also knowing the first order partial derivatives (Neumann boundary condition) for the original surface. In addition, Oliensis [137] proved that this solution is unique, whereas the theoretical number of solutions for the general Equation 2.22 is unknown. However, if the above C2 continuity assumption is violated, the surface exhibits convex/concave ambiguity as demonstrated in Figure 2.5.

There is extensive literature on SFS with the surveys by Zhang et al. [208], Durou et al. [40] summarising most classical results. Here, three classical methods are presented. The first method is by Kimmel and Bruckstein [88] and involves realising that SFS specifies the rate of change of height perpendicular to the level-sets (i.e the gradient direction which is exactly what the geometric meaning of the Eikonal equation is). Thus they pose the SFS problem as a PDE on the parametrisation of level-set curves and this allows their refinement starting from an initial estimate. Daniel and Durou [35] proposed a global error function using the irradiance equation as well as the integrability constraint and subsequently minimised it with

[3]First and second order derivatives have to be continuous .

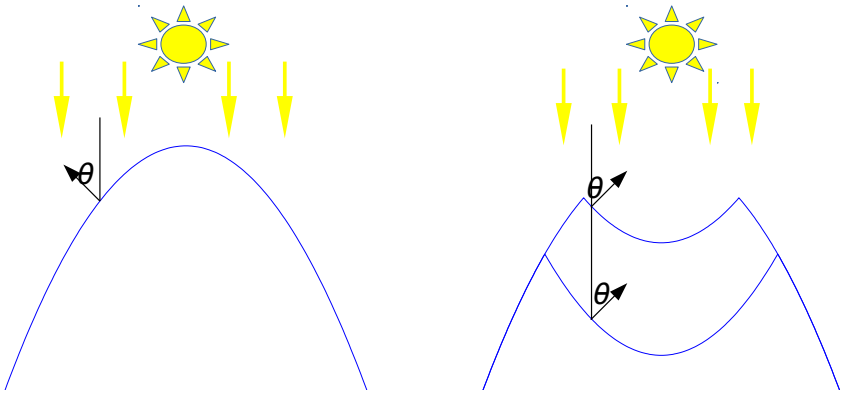


Figure 2.5 Demonstration of the convex/concave ambiguity: on the left there is a purely convex surface and on the right a family of convex/concave surfaces that will produce identical shading under vertical uniform illumination. The physical reason for this ambiguity is the fact that the SFS constraint specifies only the angle of the surface normal (marked θ here) with the lighting vector and thus there is a family of valid solutions.

gradient descent. Finally, Ping-Sing and Shah [147] linearised the non-linear denominator using a first order Taylor expansion and then subsequently linearised derivatives using finite differences to obtain a linear system. They proceeded to solve that using the Jacobi method since a direct solution was computationally impossible back then.

Finally, more recently, a number of works have fused shading information with other 3D reconstruction techniques, as the increased amount of information helps to overcome the inherent ambiguities of SFS. Jin et al. in [77, 76] proposed a multi-view SFS method which extended multi-view stereo and shape from shading, and involved a unified variational optimisation that fuses data from multiple images in a single system. In addition, other approaches have combined SFS with data from active depth sensors, including Wu et al. [198], Maier et al. [111] which refine depth maps fused into a signed distance field parametrisation (see Chapter 6) as well as Haefner et al. [54] which complements the SFS ambiguity with the depth super-resolution ambiguity in order to derive a well-posed problem.

2.4 Deep learning for Photometric Stereo

As deep learning approaches have recently become very popular in the computer vision community, dominating the vast majority of published papers and offering state-of-the-art results in a lot of problems even including geometry, they are worth mentioning. Most of the challenges in Photometric Stereo come from the difficulty of inverting non-linear and complex BRDFs (see Section 1.4.3). Therefore, deep learning seems like an obvious way to go as it is known to be able to approximate highly non-linear mappings. However, regressing

very dense and accurate depth maps is not a trivial task, although some recent approaches have had moderate success for example Eigen et al. [42], Garg et al. [50], Kendall et al. [86] . Nonetheless, these approaches have mostly aimed at very specific tasks, such as road scenes, due to availability of very detailed datasets for training and testing (e.g. KITTI [123]), and extending them so as to be able to handle arbitrary objects, is far from trivial.

There are some Deep Learning approaches for PS that are able to predict normal maps under the assumption of directional uniform lighting and no Lamberian reflection Santo et al. [158]. Also Chen et al. [27] very recently extended the approach to the uncalibrated scenario, where the directional lighting is unknown. These approaches rely on generating high quality synthetic data for training since training on real data is practically impossible due to the fact that most other 3D reconstruction methods give less accurate results than Photometric Stereo, and thus are unsuitable for generating training data or reliable ground truths (e.g see comparison with MVS and laser in Figure 1.3). However, extending these approaches to the more interesting near field scenario has not yet been achieved. One of the main difficulties is the fact that non-linear light attenuation from point light sources makes the translational invariance property of convolutional neural networks irrelevant. To overcome this limitation, the training data should probably include point light sources covering the whole 3D volume, all of which requires a much more extensive data set which also comes with more computational requirements.

Chapter 3

Theoretical framework

In this chapter, the basis of several modern Photometric Stereo approaches is presented. It consists of a differential framework where the key notion is that the surface normal can be expressed as a function of the depth gradient. This change of variables poses the PS problem as a system of PDEs with the depth being the unknown. This formulation allows for direct depth computation essentially bypassing the numerical integration of the normal field step presented in Section 2.2.2. The chapter begins with a presentation of the differential approach to the simple classic PS problem (see Section 2.2) with all the respective assumptions. The necessary steps in order to overcome these limitations and deal with increasingly realistic situations, are then discussed.

Finally, solutions to the PDEs arising through the PS problem are explained, including a simple fast marching approach (Section 3.6), a least-squares variational solver (Section 3.7) and finally an ADMM-based robust solver in Section 3.7.4.

3.1 Differential approaches to Photometric Stereo

The idea of expressing the shading problem in differential form is older than the PS problem as this was the default way to deal with the shape-from-shading situation (see Section 2.3). First of all, the differential normal parameterisation of Equation 2.15 is substituted into the classic PS image irradiance Equation 2.11 and gives (by ignoring the dependence on u, v for clarity as well as using the aliases $\frac{\partial z}{\partial u} = z_u$ and $\frac{\partial z}{\partial v} = z_v$):

$$i_j = \rho \overline{[z_u, z_v, -1]}^{\mathsf{T}} \cdot \overline{\boldsymbol{L}}_j) = \rho \frac{\bar{l}_j^1 z_u + \bar{l}_j^2 z_v - \bar{l}_j^3}{\sqrt{z_u^2 + z_v^2 + 1}}. \tag{3.1}$$

Equation 3.1 closely resembles the SFS differential Equation 2.22 with the additional unknown albedo ρ.

Solving the system of equations 3.1 (one for each image j) is not straightforward because of the non-linear dependence on the derivatives, as well as the dependence on the albedo ρ. The only method that has attempted to solve it directly has been the very recent Radow et al. [155] employing matrix differential calculus and the inertial proximal optimisation strategy (Ochs et al. [135]) . Still, the method of [155] is only applicable under the classic PS assumptions and does not trivially extend to more realistic situations. Instead, a number of methods have been proposed to simplify this non-linear PDE system and many of them are reliant on the idea of *photometric ratios*.

3.2 The image ratio method

The idea to consider image ratios first appeared in Davis and Soderblom [36], in order to eliminate the dependence on albedo on planetary images. The method of the image ratios was also used for binocular stereo matching (Wolff and Angelopoulou [191]), PS and stereo fusion (Lee and Kuo [97]) and it was also used to show that for any two images there is always an object and two lighting conditions consistent with these images (Jacobs et al. [75]). Alldrin and Kriegman [5] and Chandraker et al. [25] used photometric ratios in order to get invariants for surfaces with arbitrary reflectance. Finally, ratio-based photometric invariants were used to increase robustness in shadows and other imperfections in Wu et al. [200] and Mecca et al. [122].

Linearising the differential form of PS The importance of photometric ratios lies in their ability to simplify the differential formulation of PS (Equation 3.1) as proposed in Mecca and Durou [114] and Mecca and Falcone [115]. Indeed, they consider the ratio of Equations 3.1 for images i_j and i_k that eliminates the unknown albedo ρ as well as the square root at the denominator giving:

$$\frac{i_j}{i_k} = \frac{\bar{l}_j^1 z_u + \bar{l}_j^2 z_v - \bar{l}_j^3}{\bar{l}_k^1 z_u + \bar{l}_k^2 z_v - \bar{l}_k^3}. \tag{3.2}$$

We note that the ratio operation is mathematically valid assuming $i_j, i_k > 0$ and $\rho > 0$ and $\overline{\boldsymbol{N}}(u, v) \cdot \overline{\boldsymbol{L}}_{j,k} > 0$. This is the case for all pixels inside the image domain for which the irradiance equation is valid. This assumption breaks down at cast shadows and at points with zero reflectance $\rho = 0$. Avoiding these points is a numerical issue that is discussed in Section 3.5.1.

For now, re-arranging Equation 3.2, gives the following *linear* partial differential equation:

$$z_u(i_j\bar{l}_k^1 - i_k\bar{l}_j^1) + z_v(i_j\bar{l}_k^2 - i_k\bar{l}_j^2) = (i_j\bar{l}_k^3 - i_k\bar{l}_j^3) \tag{3.3}$$

$$\Downarrow$$

$$\boldsymbol{b} \cdot \nabla z = s \tag{3.4}$$

with

$$\boldsymbol{b} = [i_j\bar{l}_k^1 - i_k\bar{l}_j^1, i_j\bar{l}_k^2 - i_k l_j^2]^\intercal \quad , \quad s = i_j\bar{l}_k^3 - i_k\bar{l}_j^3 \tag{3.5}$$

being a vector and a scalar field only dependent on the image data. As it turns out (this is implied from Theorem 3.6.1), $\boldsymbol{b} \neq \mathbf{0}$ and thus if z is known at the boundary of the image domain, its value can be propagated to the entire domain by following the characteristic field (i.e the vector field $\boldsymbol{b}$). In addition, the need for boundary conditions is eliminated by assuming at least another image q; then another version of Equation 3.3 exists, for image pair j, q.

More details about solving this linear PDE will be presented in Section 3.6. For now, it will be discussed how the above presented approach can be extended to model more realistic effects that violate classic PS assumptions.

3.3 Near field

This section describes how the differential approach for the classic PS problem presented in Section 3.2 can be extended to model *near field* effects. Near field in this context means to assume the following two violations to the classic PS assumptions (Section 2.2): firstly, we assume that the scene is so close to the camera that the orthographic projection model becomes invalid and perspective projection applies. Secondly, and more importantly, under a near field scenario, the illumination cannot be assumed parallel and uniform; this is mostly applicable in practical systems using LEDs as light sources which result in the point light source assumption.

3.3.1 Camera model

Perspective viewing geometry for photometric 3D-reconstruction was introduced by Horn [68], then used by Bruckstein [22], and later developed in a comparable manner by Tankus and Kiryati [174], Prados and Faugeras [148] and Durou and his co-workers [20, 33]. An influential work by Papadhimitri and Favaro [142] introduced a new parameterisation in order to deal with the uncalibrated PS problem (see section 5.2.1). The reason why this

parameterisation is very useful is because it extends the perspective viewing projection to that of orthographic projection by simply making the focal length tend to infinity. More specifically, the corresponding real world 3D point (x, y, z) of the image point u, v with depth z, is given in Equation 3.6:

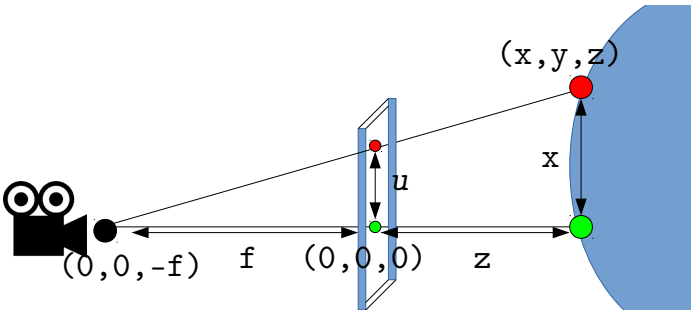


Figure 3.1 Perspective projection parameterisation as introduced by Papadhimitri and Favaro [142]. The key insight gained is placing the origin on the camera plane instead of the camera center, which had been the case in MVS literature (Harltey and Zisserman [56]). Thus the parametrisation is well defined for $\lim_{f\to\infty}$ and models the orthographic projection.

$$[u, v](x, y, z) = \Pi(x, y, z) = \left[\frac{fx}{z+f}, \frac{fy}{z+f}\right] \tag{3.6}$$

or equivalently

$$\begin{aligned} x(u, v) &= \frac{u\big(z(u, v) + f\big)}{f} \\ y(u, v) &= \frac{v\big(z(u, v) + f\big)}{f}. \end{aligned} \tag{3.7}$$

It has to be noted that this parameterisation is slightly different to the one found in Harltey and Zisserman [56], which is widely used in MVS literature. The difference lies in that the origin from the camera center is moved to its projection at the camera plane. This adds the $+f$ factor at the numerator and ensures that $\lim_{f\to\infty} x(u, v) = u$ and $\lim_{f\to\infty} y(u, v) = v$, and thus corresponds to orthographic camera parameterisation.

According to this parameterisation, the direction of the outgoing normal to the surface is given by:

$$\boldsymbol{N}(u, v) \propto \left[\nabla z(u, v), -\frac{f + z(u, v)}{f} - \frac{(u, v)\cdot\nabla z(u, v)}{f}\right]^{\mathsf{T}} \tag{3.8}$$

where the derivatives of z are considered in the image plane coordinates, and are denoted in the following $\nabla z(u, v) = \left[\frac{\partial z}{\partial u}(u, v), \frac{\partial z}{\partial v}(u, v)\right]$.

Not surprisingly, this parameterisation collapses to the standard parameterisation of the classic PS (Equation 2.15) for infinite focal length:

$$\lim_{f\to\infty} \boldsymbol{N}(u,v) \propto [\nabla z(u,v), -1]^\intercal. \tag{3.9}$$

The introduction of perspective projection makes several aspects of the PS modelling harder to handle analytically: for example $r_1 = \frac{\boldsymbol{N}^1}{\boldsymbol{N}^3} = \frac{-fz_u}{f+z+uz_u+vz_v}$ and $r_2 = \frac{\boldsymbol{N}^2}{\boldsymbol{N}^3} = \frac{-fz_v}{f+z+uz_u+vz_v}$ so the normal integration (see Section 2.2.2) is not straightforward due to the non-linear dependence on z. This specific nonlinearity can be avoided by using the substitution $g = \log(z + f)$ ([142, 120]) and so $g_x = \frac{z_u}{z+f} = \frac{-r_1}{f+ur_1+vr_2}$ and $g_v = \frac{z_v}{z+f} = \frac{-r_2}{f+ur_1+vr_2}$ are independent of the unknown g and thus standard integration methods (e.g Frankot and Chellappa [48]) can be used.

3.3.2 Nearby light sources

Starting from some initial works dealing with uniform and directional lighting assumptions, light propagation models have been extended to consider more general lighting conditions. This is very important as in practice it is rather hard to create uniform directional lighting in a lab setting; LEDs that are more closely related to ideal point light sources are commonly used and thus light attenuation needs to be modelled. This means that the further away a point is from the light source, the less light it receives, usually in an inverse square law.

A noteworthy exception to the impracticality of uniform directional illumination is a class of PS methods that perform outdoors PS, using the sun as a light source (which is far away enough for the illumination to be parallel). Shen et al. [166] showed that, contrary to common belief, the sun path in the sky actually deviates from simple planar motion and so the corresponding linear system is full rank and thus solvable. Abrams et al. [1] maximised the photometric parallax by collecting data over a period of a few weeks and Hold-Geoffroy et al. [64] observed that the sun's partial occlusion from clouds leads to a mean lighting which constantly varies, thus eliminating the need for acquisitions over long time periods.

The idea of using point light sources for PS was introduced by Iwahori et al. [74] and later adopted by Clark [30]. Iwahori et al. [74] expressed light attenuation in terms of unknown depth and proceeded to solve the non-linear equation per pixel. More recently, Migita et al. [124] proposed a non-linear optimisation framework closely resembling the "bundle adjustments" optimisation procedure for structure-from-motion problems and Papadhimitri and Favaro also extended their near field parameterisation (see Section 3.3.1) to deal with point light sources in [143]. Xie et al. [201] suggested incremental steps from fully orthographic

assumptions to a perspective near field context in order to provide accurate reconstruction according to specific mesh deformations.

An intermediate situation between point light sources and fully uniform illumination is the approach of Clark [31] using a flat LCD display which illuminates with different pre-defined patterns. Another approach is that of Hung et al. [70] who used a mirror sphere configuration to capture complex environment lighting without modelling it explicitly. Any alteration in the environment, which is not restricted to the dominant light source, can be used to produce photometric parallax. However, this method relies heavily on the assumption that all the light sources are directional.

Finally, Liu et al. [101] recently parameterised the surface with a triangle mesh, expressing everything in terms of the vertex positions using a circular ring of light sources to get to a relatively simple optimisation. In addition, they calibrated the light source positions through their reflection on a flat LCD display. Quéau et al. [149] provided an extensive survey of methods to deal with LED light sources and also proposed an alternating optimisation for the semi-calibrated PS problem (see Chapter 5 for an explanation and comparison).

Analytic modelling In order to have a near-lighting parameterisation suitable for the algebraic manipulation of equations, the approach of Mecca et al. [121] is followed. This involves modelling two effects: variable lighting direction and illumination intensity for all pixels on the image domain. To consider these effects, an important quantity is the lighting vector field $\boldsymbol{L}(u, v)$, which now depends on the surface depth and is defined in Equation 3.10, in terms of the light source position $\boldsymbol{S}_j$:

$$\boldsymbol{L}_j(u, v) = \boldsymbol{S}_j - \boldsymbol{X}(u, v). \tag{3.10}$$

More specifically, for a point $\boldsymbol{X}(u, v) = [x, y, z]^\intercal(u, v)$ in 3D space, the radiant flux is attenuated due to distance. This attenuation factor a_d is then:

$$a_d = \frac{1}{\|\boldsymbol{L}(u, v)\|^2}. \tag{3.11}$$

This radial attenuation introduces another term into the image irradiance equation and adds more non-linearities. Thus, resolving the problem requires an iterative procedure which is discussed in Section 3.7.6).

3.3.3 Angular anisotropy

It is very common for most LED light sources to exhibit angular anisotropy. This means that one "principal direction" receives maximum illumination and other directions receive less. More specifically, this attenuation factor is dependent on the principal direction of light source j, $\boldsymbol{P}_j$ and the angle between that and the lighting vector, so $\overline{\boldsymbol{P}_j} \cdot \overline{\boldsymbol{L}_j(u,v)}$. In general, the anisotropy factor a_{ang} is a general function $a_{ang}(\boldsymbol{P}_j \cdot \boldsymbol{L}_j(u,v))$ dependent on the type of the source and usually calibrated by manufacturers, under the form of luminous intensity diagrams. A good approximation for many commercial LEDs (Quéau et al. [149]) is the *imperfect Lambertian primary source model*, which equals to:

$$a_{ang}(u,v)_j = (\overline{\boldsymbol{P}_j} \cdot \overline{\boldsymbol{L}_j(u,v)})^{\mu} \tag{3.12}$$

where $\mu \geq 0$ is the anisotropy factor. Isotropic sources are obtained by setting $\mu = 0$, while stronger anisotropy effects are modelled by increasing the value of μ.

3.3.4 Overall attenuation

Finally, the intrinsic brightness of the light source is represented by a single scalar ϕ_j. This can be measured with a LUX meter or approximately computed (see Chapter 5).

Combining both attenuation factors from Equations 3.11 and 3.12, the following general attenuation factor is:

$$a_j(u,v) = \phi_j a_d a_{ang} = \phi_j \frac{(\overline{\boldsymbol{P}_j} \cdot \overline{\boldsymbol{L}_j(u,v)})^{\mu}}{\|\boldsymbol{L}_j(u,v)\|^2}. \tag{3.13}$$

3.4 Photometric Stereo for specular surfaces

Most of the PS approaches have dealt with purely diffuse reflectance (Lambertian surfaces). However no real material is purely diffuse[1] and thus this limits the applicability of the PS method. A very common strategy is to deal with specular highlights as sparse outliers and follow a robust estimation method (see Section 2.2.1). However, a number of approaches have attempted to leverage the information contained in the specular reflection instead of simply ignore it.

[1]In fact, highly diffuse commercial paints are available at a high cost, for example the barium sulphate (BaSO4) coating from http://www.pro-lite.co.uk/File/barium_sulphate_coating.php claiming 98% diffuse reflectance costs hundreds of pounds per litre.

3.4.1 Earlier works

One of the earliest approaches to PS for specular surfaces is Ikeuchi [73] who introduced a smoothness prior on the surface along with the use of a lookup table, thereby limiting the approach to some specific industrial applications. Nayar et al. [128] modelled specular reflection with a Dirac delta function and used a spherical diffuser to get an extended light source setup in order for the irradiance equation to have a simple form. Wolff [190] proposed a reflectance model for combined diffuse and specular reflections from dielectric materials. Since the modelling involved purely physical parameters in terms of the Fresnel coefficients and the Dirac delta function, the results achieved can be considered mostly theoretical rather than actually being applicable for shape reconstruction. Goldman et al. [53] described the irradiance equation as a weighted sum of fundamental diffuse and specular materials (using the Ward [186] model), and solved it iteratively using non-linear least squares. Georghiades [51] proposed an iterative optimisation method over a very large set of variables, using the Torrance and Sparrow reflectance model for the uncalibrated PS problem (see Section 5.2.1). Chung and Jia [29] adopted a two-step optimisation to estimate the normals and all reflectance parameters. In the first step, occluding boundaries (which had been detected through shadows) were used to find points with a known surface normal. These recovered normals played a key role in estimating the global reflectance parameters according to the Ward BRDF . In the second step, with the reduced number of unknowns, a robust iterative optimisation process was applied, estimating the surface normals and the diffuse albedo for all surface points.

A special mention goes to a few methods that deal with specular surfaces indirectly. Johnson et al. [80, 81] developed an elastomer that is attached to the surface to change its BRDF to Lambertian so standard PS is applicable. Hertzmann and Seitz [61] removed the need for an explicit BRDF model by comparing object images with images of reference objects. Thus the PS problem is posed as a matching problem and machinery from the binocular stereo literature can be used. Barsky and Petrou [10] detected and removed specular highlights by observing differences in the chromaticity. Alldrin and Kriegman [5], overcame standard reflectance assumptions by exploiting the bilateral symmetry of real BRDFs ("Helmholtz reciprocity", see Section 2.1) in a specific lighting setup. Chandraker et al. [24] followed a similar procedure to get photometric invariants from temporal derivatives for a surface with an isotropic BRDF and light source movement on a circle. Hyeongwoo et al. [71] separated diffuse and specular components of images using the "Dark channel" of the image (defined as the minimum of all colour channels at the neighbourhood of a pixel). Finally, Yang and Ahuja [202] introduced a PS method for estimating diffuse reflection and surface normals from color images. Using a dichromatic reflection model, they considered surface chromaticity as

a matching invariant for PS assuming that at least one of the color images corresponded to diffuse reflection.

3.4.2 Unifying diffuse and specular reflection in a single term BRDF

Despite the extensive literature that deals with specular surfaces, most methods described above are not suitable to embed in a differential parameterisation and be analytically handled. As seen in section 2.1, there are multiple analytical light reflection models which can approximate real BRDFs and be handled analytically. In addition, single term BRDFs are highly desirable as this allows for the ratio method (Section 3.2) to be used. Thus, Mecca and Quéau [117] introduced the following BRDF, aiming to unify Lambertian and Blinn-Phong in a single term:

$$i = \left(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}(\boldsymbol{L}, \boldsymbol{V})\right)^{\frac{1}{c}}. \tag{3.14}$$

In Equation 3.14, $\boldsymbol{W}$ is the *weighted* half vector defined as:

$$\boldsymbol{W} = \overline{\boldsymbol{L}} + \min\left\{1, \frac{|1-c|}{\varepsilon}\right\}\overline{\boldsymbol{V}}. \tag{3.15}$$

Here ε is a parameter describing the material of the surface, allowing the combination of diffuse and specular reflectance by taking into account the continuous transition state between the diffuse and specular component. Equation 3.14 can be used to deal with surfaces made by heterogeneous materials depending on c and ε:

- by choosing $c = 1$, $\boldsymbol{W} = \boldsymbol{L}$ and so we obtain purely Lambertian reflection;
- by setting $0 < c \leq 1 - \varepsilon$, $\boldsymbol{W} = \boldsymbol{H}$ leading to purely specular Blinn-Phong reflectance, where the size of the specular lobes depend on the value of c;
- the choice $1 - \varepsilon < c < 1$ provides a transition state between the diffuse and specular component;
- spatially-varying reflectances can be modelled by using the space dependency of c.

The most similar approach to that presented here is that of Tozza et al. [177] as they write a PDE, mixing the PDEs generated using the ratio method with diffuse and Blinn-Phong specular terms.

Despite being empirical, the reflectance model presented above unifies diffuse and specular reflection in a single term, thus making the ratio method still applicable (as explained in depth in Section 3.5). However, the subsequent question that naturally arises is the ability

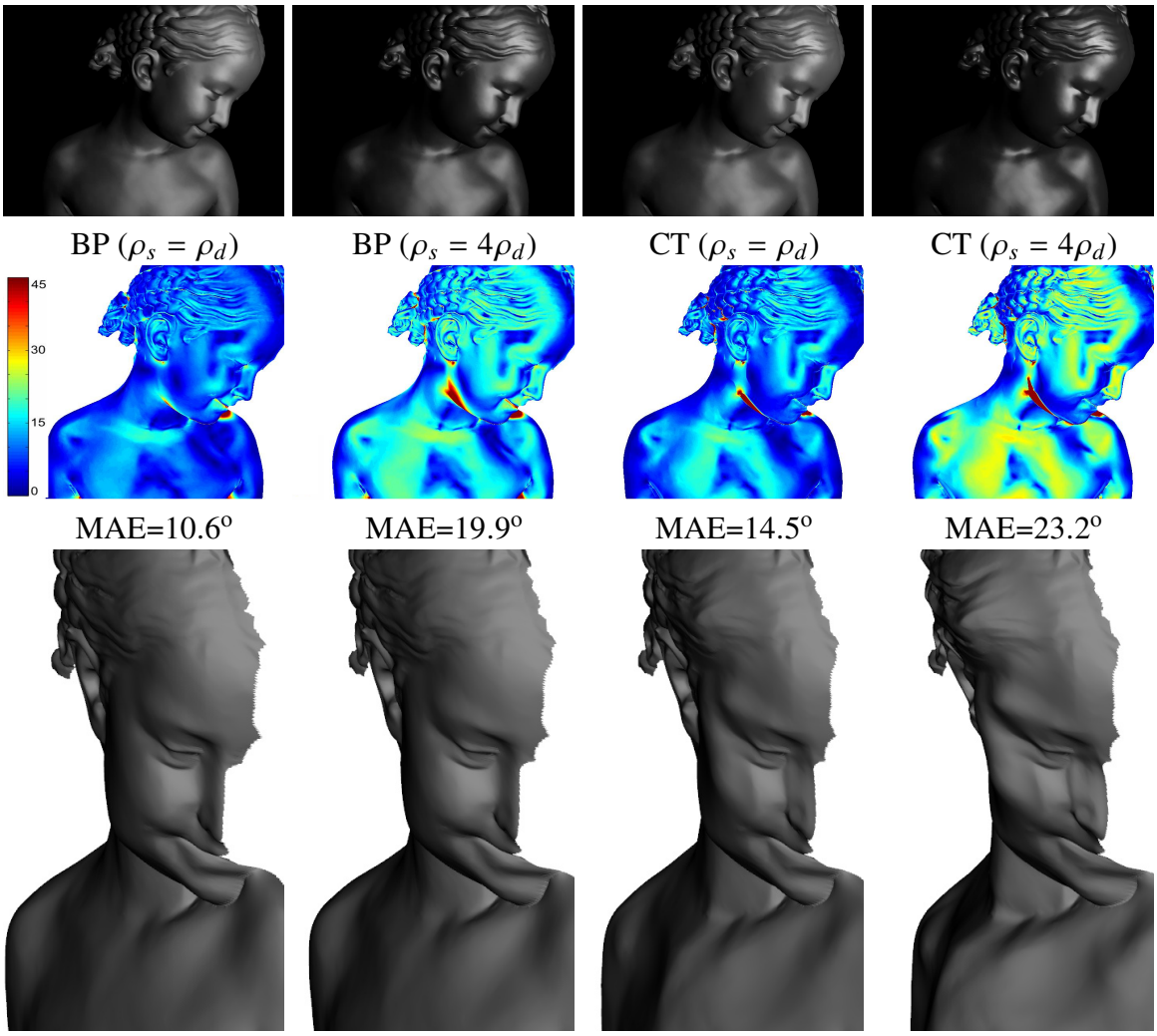


Figure 3.2 Evaluation of the proposed method (based on the irradiance Equation 3.14) on synthetic data rendered with various reflection models. Top row: 1 sample image for each dataset. The specular reflection model is Blinn–Phong (BP) for two datasets on the left and Cook & Torrence (CT) for the datasets on the right (see Section 2.1.2 for reflection models). Middle row: normal error maps compared to the ground truth (in degrees). Bottom row: 3D reconstructions. It is noted that the proposed irradiance model performs adequately at data rendered with more standard (in graphics literature) reflection models.

of this BRDF to model more complex reflections. In Mecca et al. [118] this was verified experimentally as shown in Figure 3.2. These experiments consider synthetic images rendered with a 2 term irradiance equation $i = \rho_s i_s + \rho_d i_d$, where ρ_s and ρ_d are specular and diffuse albedos respectively (and $\rho_s + \rho_d = 1$) and i_s and i_d are specular and diffuse reflection components. For i_s, two cases are considered: standard Blinn-Phong and Cook & Torrence (for details see Section 2.1.2). As seen in Figure 3.2, the proposed BRDF performs adequately

for data rendered with two term BRDFs, as well as for data created using the more realistic Cook and Torrence BRDF.

3.5 The unified equation

Having discussed modelling strategies for near-field effects in Section 3.4 and non-Lambertian reflection in Section 3.4, this section uses the image ratio method explained in Section 3.2 in order to derive a PDE describing the PS problem. Therefore, the overall irradiance equation for image j is:

$$i_j(u,v) = \rho(u,v)a_j(u,v,z(u,v))\left(\overline{\boldsymbol{N}}(u,v,z(u,v))\cdot\overline{\boldsymbol{W}}_j(u,v,z(u,v),c)\right)^{\frac{1}{c(u,v)}} \tag{3.16}$$

Raising both sides of Equation 3.14 to the power of c and then considering the ratio of equations for images j and k gives:

$$\frac{i_j(u,v)^{c(u,v)}}{i_k(u,v)^{c(u,v)}} = \frac{a_j(u,v,z(u,v))^{c(x,y)}\boldsymbol{N}(u,v,z(u,v))\cdot\overline{\boldsymbol{W}}_j(u,v,z(u,v),c)}{a_k(u,v,z(u,v))^{c(u,v)}\boldsymbol{N}(u,v,z(u,v))\cdot\overline{\boldsymbol{W}}_k(u,v,z(u,v),c)}. \tag{3.17}$$

Expanding out Equation 3.17 and substituting the parameterisation of the normal from Equation 3.8, after some algebra, gives the following equation:

$$\boldsymbol{b}_{jk}(u,v,z(u,v),c(u,v))\cdot\nabla z(u,v) = s_{jk}(u,v,z(u,v),c(u,v)). \tag{3.18}$$

where (ignoring some dependencies for better readability):

$$\begin{aligned}\boldsymbol{b}_{jk} = \Big[&\left(a_k i_j\right)^c\left(\overline{W}_k^1 - \frac{u}{f}\overline{W}_k^3\right) - \left(a_j i_k\right)^c\left(\overline{W}_j^1 - \frac{u}{f}\overline{W}_j^3\right),\\ &\left(a_k i_j\right)^c\left(\overline{W}_k^2 - \frac{v}{f}\overline{W}_k^3\right) - \left(a_j i_k\right)^c\left(\overline{W}_j^2 - \frac{v}{f}\overline{W}_j^3\right)\Big]^\intercal\end{aligned} \tag{3.19}$$

and

$$s_{jk} = \frac{f+z}{f}\left(\left(a_k i_j\right)^c\overline{W}_k^3 - \left(a_j i_k\right)^c\overline{W}_j^3\right). \tag{3.20}$$

Equation 3.18 will be referred to from now on as the unified equation. This is a quasi-linear, first order PDE, meaning that it is linear in terms of the gradient of ∇z but it has a non-linear dependence on z, mostly because of the light propagation and attenuation terms. We note that this is a generalisation of Equation 3.3.

Colored scenes and materials As remarked by Quéau et al. [152], color images are automatically handled with the framework presented above: since each color channel provides one unifying equation, they can be stacked in the same way as above. This is another remarkable result since there is no need to assume that either the source or the material (e.g. Brostow et al. [21]) is monochromatic or piece-wise monochromatic (e.g. Anderson et al. [6]). This assumption is usually introduced because it is impossible to separate the colored light intensity from the colored albedo. Since this formulation is independent of the albedo, only the directions and intensities of the lights have to be explicitly considered.

3.5.1 Missing data

If for some pixels and in some images the image data are clearly unreliable, for example because of the pixel value being saturated or completely black (due to the quantisation of image data, usually in 10 bits, any irradiance value smaller than $1/1024$ of the maximum will lead to a 0 pixel value), the following equations can be ignored in order to minimise the potential number of outliers. In addition, if for some point all the data are missing (for example because of zero albedo which means that no light is reflected), we can assume that the surface is flat, i.e. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \nabla z = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. This post-processing step of the calculated $\boldsymbol{b}_{jk}$, s_{jk} fields will perform "hole-filling" in this zone and ensure numerical stability. Note that this procedure is conceptually similar to a regularisation step proposed by Hernández et al. [60] in order to overcome regions with shadows (that act as missing data).

3.5.2 Image ratio criticism

After having introduced the image ratio method and having described how it can be used to simplify the PS formulation to lead to an albedo independent quasi-linear PDE, it is worth noting its main drawbacks. That is that the algebraic error in the equations has no physical meaning, as it does not directly map to an error in the image irradiance equation.

For someone familiar with the SFM literature, this may seem like a very big problem. Indeed in [56], minimising the image re-projection error for solving the SFM problem is referred to as "The Gold Standard algorithm" and the literature almost universally solves "bundle-adjustment" type problems, where camera positions and points are optimised through non-linear minimisation of the re-projection error (for recent results see Hong [65]).

However, the situation is different in PS, as image irradiance equations do not really model the real image formation process in the same level as projection of SFM points. To make matters worse, virtually every piece of literature including this thesis assumes a

simplified lighting model (see Section 2.1.2), and thus the importance of the re-projection error is minimised. In addition, virtually every realistic scene would have some outlier points (shadows, depth discontinuities), which degrade the importance of the re-projection error even more. In sort, even a ground truth surface reconstruction will not have a zero re-projection error.

Another theoretical disadvantage of the ratio method is that it could potentially lead to numerical instabilities. Indeed, as ratios are numerically unstable for very small numbers, pixels where the image value is small should be avoided as described in Section 3.5.1. This is somewhat alleviated by the fact that pixel value ratios are never actually computed, instead the $\boldsymbol{b}$ is calculated by expanding out the relevant equations, and thus is proportional to image intensity i. Therefore in a very low albedo region with $i_j \approx i_k \approx 0$, the fundamental equation reduces to $0 = 0$ which is meaningless but unlikely to bias the solution in other, better constrained, data points.

Finally, it is noted that the ratio method drastically alters the noise distribution. For example, even if the noise in the original images is assumed to follows a Gaussian distribution (which is never really true), their ratio follows a Cauchy distribution (Pham-Gia et al. [145]), which is much harder to do exact inference for. This in practice means that least-square based estimators are expected to be suboptimal, stressing out the need for more robust loss functions (see Section 3.7.4).

3.6 Solving the unified equation

In this section an approach for solving the PS problem expressed by the quasi-linear PDE 3.18 is presented. First of all, two important theorems (see Appendix A.1 for the proofs), are presented which establish the fact that the problem is well-posed and the solution can be obtained through a fast-marching scheme.

Theorem 3.6.1 *Let* $\mathbf{b}_{jk}(u, v, z(u, v))$ *be the vector field of (3.19) for image pair* (j, k)*. Then:*

$$\|\mathbf{b}_{jk}(u, v, z(u, v))\| > 0. \tag{3.21}$$

Theorem 3.6.1 guarantees that the differential formulation is meaningful at all points in the domain. In addition, it is also true that two different pairs of images in a general tree light source setting lead to linearly independent $\boldsymbol{b}$ fields:

Theorem 3.6.2 *For light sources j,k and q, such as rank*$([\mathbf{W}_j, \mathbf{W}_k, \mathbf{W}_q]) = 3$*, the vector fields* $\mathbf{b}_{jk}(u, v, z(u, v))$ *and* $\mathbf{b}_{jq}(u, v, z(u, v))$ *are linearly independent.*

In Theorem 3.6.2, the rank constraint condition is easily satisfied in practice for all surface points in front of the capture setup of Figure 1.5. An example of a corner case is a point with Lambertian reflectance (hence $c = 1$ and thus $\mathbf{W} = \mathbf{L}$) that is on the same plane as the light sources; however this is completely impractical as a point like that would be out of the field of view of the camera.

The importance of Theorem 3.6.2 is that it establishes a fast-marching propagation scheme which is the natural/naive way of numerically solving the system of PDEs.

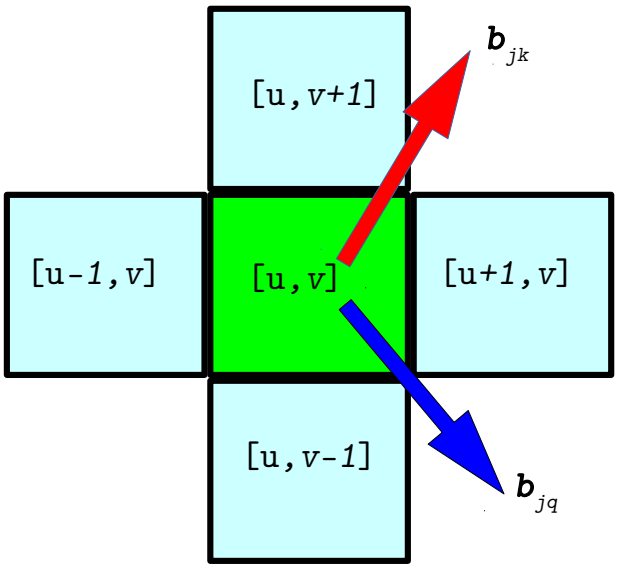


Figure 3.3 Intuitive representation of the naive fast marching way of solving the unified equation. For different pairs of images, the respective $\boldsymbol{b}$ vectors are linearly independent and thus span the 2D space (marked red and blue here). Thus the z value can be propagated from a single point to its neighbours and applying this procedure recursively fills the whole image domain.

We first start with the simple case of classic PS assumptions thus the unified Equation 3.23 collapses to Equation 3.3, i.e. the vector field $\boldsymbol{b}$ and scalar field s are independent of z and can be calculated from image data (using the known lighting vectors). To proceed with a solution on a *pixel grid*, the gradient operator is discretised as:

$$\nabla z(u,v) = \begin{bmatrix} \frac{\partial z(u,v)}{\partial u} \\ \frac{\partial z(u,v)}{\partial v} \end{bmatrix} = \begin{bmatrix} z(u+1,v) - z(u,v) \\ z(u,v+1) - z(u,v) \end{bmatrix} \tag{3.22}$$

where the first order forward differences are used and the distance between neighbouring pixels is 1. Note that for points in the boundary which are missing forward neighbours, backward differences are required but the linear form is essentially the same. Thus, substituting into

Equation 3.18 leads to:

$$b^1_{jk}(u,v)\Big(z(u+1,v)-z(u,v)\Big)+b^2_{jk}(u,v)\Big(z(u,v+1)-z(u,v)\Big)=s_{jk}(u,v) \tag{3.23}$$

$$\Downarrow$$

$$b^1_{jk}z(u+1,v)+b^2_{jk}z(u,v+1)=s_{jk}+z(u,v)\Big(b^1_{jk}+b^2_{jk}\Big) \tag{3.24}$$

where in Equation 3.24 the dependency of b^1_{jk}, b^2_{jk} and s_{jk} on (u,v) is ignored for readability. It is important to stress that this first order approximation is dependent on continuity for $\boldsymbol{b}$ and s. A slightly more accurate approximation can be obtained by linearly interpolating $\boldsymbol{b}$, s for the differences. In other words, in Equation 3.24 set:

$$b^1_{jk}=\frac{b^1_{jk}(u+1,v)+b^1_{jk}(u,v)}{2} \tag{3.25}$$

$$b^2_{jk}=\frac{b^2_{jk}(u,v+1)+b^2_{jk}(u,v)}{2} \tag{3.26}$$

$$s_{jk}=\frac{s_{jk}(u+1,v)+s_{jk}(u,v+1)+s_{jk}(u,v)}{3}. \tag{3.27}$$

The fast marching update formula arises if we consider another pair of images (which can be j,q in the theoretical minimum case of 3 images) and then stack them into a system:

$$\begin{bmatrix} b^1_{jk} b^2_{jk} \\ b^1_{jq} b^2_{jq} \end{bmatrix} \begin{bmatrix} z(u+1,v) \\ z(u,v+1) \end{bmatrix} = \begin{bmatrix} s_{jk}+z(u,v)\Big(b^1_{jk}+b^2_{jk}\Big) \\ s_{jq}+z(u,v)\Big(b^1_{jq}+b^2_{jq}\Big) \end{bmatrix}. \tag{3.28}$$

As Theorem 3.6.2 guarantees, the system matrix is invertible and thus $z(u+1,v)$, $z(u,v+1)$ can be determined from $z(u,v)$. Thus using a similar procedure with backwards finite differences, z can be propagated from a single point (u_0, v_0) to the entirety of a connected domain, as illustrated in Figure 3.3. This leaves a single parameter ambiguity for the solution (the value $z(u_0, v_0)$) which is essentially an "integration constant". In case $p > 2$ images, one could stack all $\binom{p}{2}$ image pairs of equations and perform a least squares solution in Equation 3.28. As this would be computationally expensive, Mecca et al. [119] use the following heuristic: they select images with the highest grayscale value at pixel.

Near field Assuming nearby light sources, the unified PDE is quasi-linear and so $\boldsymbol{b}$ non-linearly depends on z making the interpolation of Equations 3.25 to 3.27 impossible to compute. Thus, the solution has to use the update formula from Equations 3.24; therefore the scheme is absolutely critically dependent on continuity in z. Thus, the propagation is expected

to perform very poorly around occlusion boundaries and this is verified experimentally in Figure 3.4.

Specular reflection We note that the procedure explained above is applicable for any value of specularity parameter c, which can also vary in the image domain i.e. $c(u, v)$. However, this propagation scheme provides no trivial way of estimating this parameter. This limitation is resolved by the usage of a variational optimisation presented in the next section.

3.7 Variational optimisation framework

The main limitation of the fast-marching approach described above is that the flow of information from one point to the rest of the domain is sensitive to noise. In addition, a severe outlier such as a shadow, can cause systematic error propagation. The way to overcome this issue is to use a variational optimisation which optimises z for all pixels simultaneously. This idea dates back to Horn and Brooks [66]. There is a great deal of theoretical work on the calculus of variation with the most relevant results in Mecca et al. [118]. However, it is important to mention that to tackle this differential formulation, at the end of the day the problem has to be discretised and solved on a pixel grid. Thus a *discretise first and then solve* strategy is used. Indeed as noted in several works (e.g. Smith and Fang [169] and Bähr et al. [8]), clever discretisation of the derivative operations can lead to a global linear equation. The neighbouring information between different pixels is encoded as a sparse neighbouring matrix. The solution of this variation problem is analogous to propagating information from all points to all directions at the same time. Hence the solution is much more robust to noise and outliers as shown in Figure 3.4. This experiment considers the half CT, half diffuse dataset (Column 3 in Figure 3.2) and then adds Gaussian noise or synthetic outliers. The outliers are small black rectangles placed in different positions in different images in order to avoid complete lack of information. Six experiments are considered with three levels of noise and three different sizes for outliers. As expected, the proposed variational approach outperforms the fast-marching scheme of Mecca et al. [119], as well as the robust local method of Ikehata et al. [72]. In addition, the error propagation for [119] is very apparent in the depth discontinuity at the bottom of the face, as well as around the synthetic outliers.

3.7.1 Discretisation

This section describes the discretisation strategy that can approximate derivatives on any connected domain (not required to be rectangular) using first order forward or backward

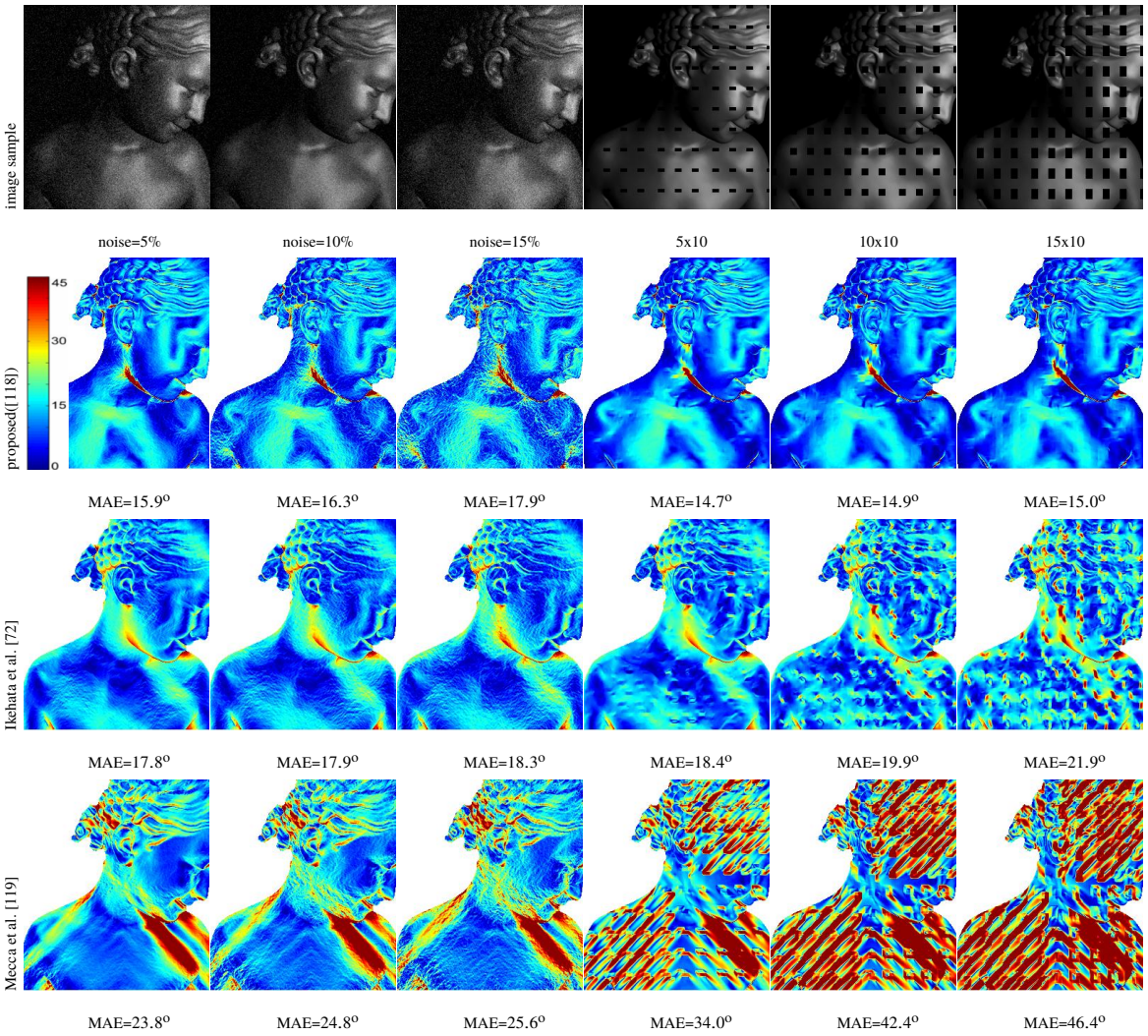


Figure 3.4 Comparison of the proposed variational solver (introduced in Mecca et al. [118]) with the fast-marching approach of Mecca et al. [119] as well as the robust *local* method of Ikehata et al. [72]. The experiments include synthetic images rendered with the Cook & Torrance model (third column of Figure 3.2) with the addition of Gaussian noise or outliers. Each column corresponds to a different experiment with the first row showing a typical image sample per experiment. Rows 2-4 show the evaluation of these methods on these experiments as a normal error map compared with the ground truth (in degrees, error bar at top left). It is observed that the proposed variational optimisation is outperforming all other approaches in all experiments. In addition, the error of the proposed method is barely increasing with the level of noise or outliers and this fact demonstrates its robustness.

differences. More precisely, if $\boldsymbol{Z}$ is the vector of all the depth values in the image domain, then:

$$\frac{\partial \boldsymbol{Z}}{\partial u} = G_u \boldsymbol{Z} \quad , \quad \frac{\partial \boldsymbol{Z}}{\partial v} = G_v \boldsymbol{Z} \tag{3.29}$$

where G_u and G_v are sparse matrices with -1 and 1 on the non-zero entries. This is similar to the discretisation in Equation 3.22 but is now applied to the whole image domain at the same time and expressed in matrix-vector form. The sparsity pattern of G_u and G_v is better explained by looking at Figure 3.5

	1	2	
3	4	5	
	6	7	8

(a) Simple domain.

$$G_u = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad G_v = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b) Gradient Matrices

Figure 3.5 Discretisation of a simple domain (a) and the respective gradient matrices (b).

Figure 3.5 shows a simple domain and the respective gradient matrices. Points on the boundary may have zero for one of the derivatives. To better express the sparsity pattern of these gradient matrices, the following notation is used: a sparse matrix A_s is written as triples of row, column, value, i.e $A_s = \text{sparse}(\boldsymbol{R}, \boldsymbol{C}, \boldsymbol{V})$. Defining $\boldsymbol{I}_c = [1, \ldots, p]^\intercal$ the index vector for rows, $\boldsymbol{I}_{up}$ and[2] $\boldsymbol{I}_{um}$ and $\boldsymbol{I}_{vp}$ and $\boldsymbol{I}_{vm}$ the index for plus and minus terms for u and v derivatives respectively, then $G_u = \text{sparse}\left(\begin{bmatrix}\boldsymbol{I}_c \\ \boldsymbol{I}_c\end{bmatrix}, \begin{bmatrix}\boldsymbol{I}_{up} \\ \boldsymbol{I}_{um}\end{bmatrix}, \begin{bmatrix}\mathbf{1}_p \\ -\mathbf{1}_p\end{bmatrix}\right)$ and $G_v = \text{sparse}\left(\begin{bmatrix}\boldsymbol{I}_c \\ \boldsymbol{I}_c\end{bmatrix}, \begin{bmatrix}\boldsymbol{I}_{vp} \\ \boldsymbol{I}_{vm}\end{bmatrix}, \begin{bmatrix}\mathbf{1}_p \\ -\mathbf{1}_P\end{bmatrix}\right)$ where $\mathbf{1}_p$ is a vector of size p containing just ones.

Thus, the unified equation $\boldsymbol{b}(u, v, z) \cdot \nabla z(u, v) = s(u, v, z) \leftrightarrow \boldsymbol{b}_u(u, v, z) z_u(u, v) + \boldsymbol{b}_v(u, v, z) z_v(u, v) = s(u, v, z)$ is written in discrete matrix form as:

$$(B_u G_u + B_v G_v)\boldsymbol{Z} = \boldsymbol{S} \text{ , where } B_u = \text{diag}(\boldsymbol{B}_u) \text{ and } B_v = \text{diag}(\boldsymbol{B}_v) \tag{3.30}$$

where $\boldsymbol{B}_u$ and $\boldsymbol{B}_v$ are the vectors containing the first and second components of the vector field $\boldsymbol{b}(u, v) = [\boldsymbol{b}_u, \boldsymbol{b}_v]^\intercal$ for all pixels (u, v) in the image domain. Simply to avoid confusion with the notation, it is emphasised that subscripts u,v denote partial derivatives for the scalar quantity z and components for vector quantities (vectors are marked in bold, no vector derivatives are defined anywhere in the thesis).

Finally, the n versions of Equation 3.30 are stacked making the combined system:

$$\left(\begin{bmatrix} B_{u1} \\ \vdots \\ B_{un} \end{bmatrix} G_u + \begin{bmatrix} B_{v1} \\ \vdots \\ B_{vn} \end{bmatrix} G_v \right) \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{S}_1 \\ \vdots \\ \boldsymbol{S}_n \end{bmatrix} \tag{3.31}$$

$$\text{i.e.} \quad A_s \boldsymbol{Z} = \boldsymbol{B}_s \tag{3.32}$$

[2]For the simple example of Figure 3.5, $\boldsymbol{I}_{up} = [2, 2, 4, 5, 5, 7, 8, 8]^\intercal$

For a number of pixels p inside the image domain and n equations, Equation 3.31 is $pn \times n$ with around $4pn$ non-zero elements in the very sparse system matrix.

3.7.2 Least squares solution

The simplest way of solving the over-constrained linear Equation 3.32 is to minimise the ℓ_2 error, i.e. $\min_{Z} \|A_s \boldsymbol{Z} - \boldsymbol{B}_s\|_{\ell^2}$ .This is found by computing the normal equations which requires multiplying both sides by the transpose of the system matrix $A_s^\intercal$. This is easily calculated by noting that all B_* matrices are diagonal and therefore equal to their transpose and thus commute under multiplication, hence:

$$\left(\left(\sum_{j=1}^{n} B_{uj}^2\right) G_u^\intercal G_u + \left(\sum_{j=1}^{n} B_{uj} B_{vj}\right)(G_u^\intercal G_v + G_v^\intercal G_u) + \left(\sum_{j=1}^{n} B_{vj}^2\right) G_v^\intercal G_v\right) \boldsymbol{Z} = G_u^\intercal \sum_{j=1}^{n} B_{uj} \boldsymbol{S}_j + G_v^\intercal \sum_{j=1}^{n} B_{vj} \boldsymbol{S}_j. \tag{3.33}$$

Despite the seemingly complicated analytical form of Equation 3.33, because of the diagonal matrices inside the summations, only the quantities $\sum \boldsymbol{b}_u^2(u,v)$, $\sum \boldsymbol{b}_u(u,v)\boldsymbol{b}_v(u,v)$, $\sum \boldsymbol{b}_v^2(u,v)$, $\sum s(u,v)\boldsymbol{b}_u(u,v)$ and $\sum s(u,v)\boldsymbol{b}_v(u,v)$ need to be calculated per pixel (with the summations occurring for all equations 1 to n) and then assigned to the indices determined by the sparsity pattern of gradient matrices.

More specifically, $G_u^\intercal = \text{sparse}\left(\begin{bmatrix} I_{up} \\ I_{um} \end{bmatrix}, \begin{bmatrix} I_c \\ I_c \end{bmatrix}, \begin{bmatrix} \mathbf{1}_p \\ -\mathbf{1}_p \end{bmatrix}\right)$ and thus:

$$G_u^\intercal G_u = \text{sparse}\left(\begin{bmatrix} I_{up} \\ I_{up} \\ I_{um} \\ I_{um} \end{bmatrix}, \begin{bmatrix} I_{up} \\ I_{um} \\ I_{up} \\ I_{up} \end{bmatrix}, \begin{bmatrix} \mathbf{1}_p \\ -\mathbf{1}_p \\ -\mathbf{1}_p \\ \mathbf{1}_p \end{bmatrix}\right) \tag{3.34}$$

and the rest of the combinations are done in the same way. Hence the sparsity pattern of the system matrix in Equation 3.33 is very predictable and so the system can be rapidly calculated with no need for matrix multiplications. This is very important in situations where there is a very large number of equations n and even computing the full set of fields $\boldsymbol{b}$ and s would consume too much memory. Instead, what needs to be done, is sequentially compute $\boldsymbol{b}$ and s and only accumulate in memory the above mentioned sums. This property of the ℓ_2 solution to drastically reduce the memory requirements will prove crucial in the multi-view solver extension (Section 6.5), as well as in the GPU implementation application (Section 7.2.1).

One remark about this simple ℓ_2 solution is the definition of $\boldsymbol{b}$ and s in Equations 3.19 and 3.20 (the quantities of the unified PDE) implicitly leads to a weighted least squares solution. The magnitude of $\boldsymbol{b}$ is higher for pairs of images and pixels with high photometric parallax (difference in $\boldsymbol{N} \cdot \boldsymbol{W}$ in two images) as well as high pixel values, although the exact dependence is complicated. Therefore, equations which are probably more numerically stable get a higher weighting, thus this is technically a weighted least squares problem.

3.7.3 Regularisation

The fast-marching approach for solving the unified PDE presented in Section 3.6 is critically dependent on knowledge of z on a single point. This is especially true for the case of directional lighting and orthographic camera projection, as the PDE only depends on ∇z and not z, so there is an implicit one parameter ambiguity. This ambiguity manifests as a rank deficient linear system in Equation 3.33, as the system matrix is expected to be positive-semi-definite with a maximum rank of $p - 1$. The naive approach would be to specify the depth at a single point (to retain the similarity with the fast marching scheme of section 3.6) by adding the respective diagonal entry. However, in practice it is possible to have even lower rank, for example in the case of a non-connected domain, which requires, in theory, at least one known point per connected region of the domain.

The state-of-the-art strategy for overcoming this limitation is to use a Tikhonov regularisation term introduced in Tikhonov and Arsenin [175]. This corresponds to augmenting an additional term $\lambda \|\boldsymbol{Z} - \boldsymbol{Z}_0\|^2_{\ell_2}$ to the ill-constrained problem $A_s \boldsymbol{Z} = \boldsymbol{B}_s$ in Equation 3.32. The quantity $\boldsymbol{Z}_0$ is a prior for the depth and can be set to a constant (i.e. the mean depth) if no better estimate is available (for example from MVS[3]).

Therefore, the overall weighted least squares solution is:

$$(A_s^\intercal A_s + \lambda I_{p \times p}) \boldsymbol{Z} = A_s^\intercal \boldsymbol{S} + \lambda \boldsymbol{Z_0} \tag{3.35}$$

where $Ip \times p$ is the identity matrix. λ controls the weight of the prior and should be relatively small (e.g $10^{-5} - 10^{-9}$) to avoid biasing the solution in case of a flat prior.

Finally, it is important to point out that this regularisation scheme also excludes the degenerate solution $z(u, v) = \infty, \ \forall (u, v)$. Indeed, infinity z leads to $a(u, v) = 0$ and thus Equation 3.18 reads as $\boldsymbol{0} \cdot \nabla z = 0$ which is exactly satisfied as $\nabla z = \boldsymbol{0}$. Not surprisingly, the regularised Equation 3.35 does not suffer from this problem.

[3] In section 6.6, an automatic strategy for overcoming this one parameter scale ambiguity will be discussed.

3.7.4 Robust loss function

In order to enhance robustness to non-Gaussian noise (see Section 3.5.2) and outliers such as shadows or non-differentiable elements of the surface, it is required that a more robust loss function instead of the simple ℓ_2 is considered. It is common in literature to assume that these outliers are sparse phenomena (see Section 2.2.1) and rely on ℓ_1 optimisation [72, 150]. Contrary to the work by Ikehata et al. [72], a global ℓ_1 optimization is considered, over the whole image domain Ω and all image ratios (j, k). Thus, we solve the discrete PS problem of Equation 3.32 in the ℓ_1 sense, by considering the following optimization problem (again including the Tikhonov regularisation term):

$$\min_{z} ||A_s \boldsymbol{Z} - \boldsymbol{B}_s||_{\ell_1} + \lambda \|\boldsymbol{Z} - \boldsymbol{Z}_0\|_{\ell_2}^2. \tag{3.36}$$

Solving Equation 3.36 is non-trivial. One way to do this is by using the Alternating Direction Method of Multipliers (ADMM) (Boyd and Vandenberghe [19]). This re-writes Equation 3.36 (using the auxiliary variable $\boldsymbol{G}$) as:

$$\begin{cases} \min\limits_{\boldsymbol{Z},\boldsymbol{G}} \|\boldsymbol{G}\|_{\ell_1} + \lambda \|\boldsymbol{Z} - \boldsymbol{Z}_0\|_{\ell_2}^2 \\ \text{s.t.} \quad \boldsymbol{G} = A_s \boldsymbol{Z} - \boldsymbol{B}_s \end{cases} . \tag{3.37}$$

Introducing the scaled dual variable[4] $\boldsymbol{U}$, problem 3.37 can be solved using the alternating direction scheme (t is the iteration number):

$$\boldsymbol{G}^{t+1} = \operatorname*{argmin}_{\boldsymbol{G}} \|\boldsymbol{G}\|_{\ell_1} + \alpha \|\boldsymbol{G} - \left(A_s \boldsymbol{Z}^t - \boldsymbol{B}_s - \boldsymbol{U}^t\right)\|_{\ell_2}^2 \tag{3.38}$$

$$\boldsymbol{Z}^{t+1} = \operatorname*{argmin}_{\boldsymbol{Z}} \frac{\lambda}{\alpha} \|\boldsymbol{Z} - \boldsymbol{Z}_0\|_{\ell_2} + \|A_s \boldsymbol{Z} - \left(\boldsymbol{G}^{t+1} + \boldsymbol{B}_s + \boldsymbol{U}^t\right)\|_{\ell_2}^2 \tag{3.39}$$

$$\boldsymbol{U}^{t+1} = \boldsymbol{U}^t + \boldsymbol{G}^{t+1} - \|A_s \boldsymbol{Z}^{t+1} - \boldsymbol{B}_s\|_{\ell_1} \tag{3.40}$$

starting from $(\boldsymbol{Z}^0, \boldsymbol{G}^0, \boldsymbol{U}^0) = (\boldsymbol{Z}_0, A_s \boldsymbol{Z}_0 - \boldsymbol{B}_s, \boldsymbol{0})$. This scheme can be proven to converge from almost any descent parameter α, whose choice only affects the convergence rate. Equation 3.39 is just a least squares $\boldsymbol{Z}$ estimation (using $\frac{\lambda}{\alpha}$ as a regularisation weight) and can be solved as described in section 3.7.2.

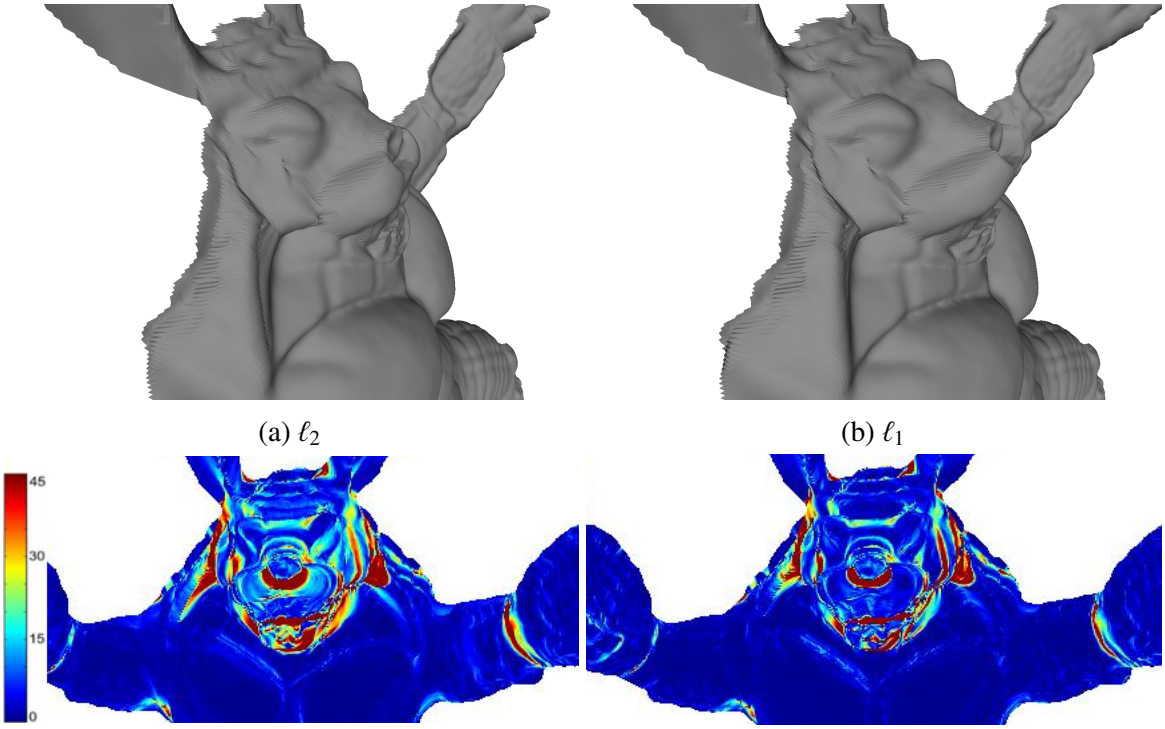
Equation 3.38 actually has a closed form solution described in the following Lemma[5].

Lemma 3.7.1 *If* $\boldsymbol{X}^* = \min\limits_{\boldsymbol{X}} (\|\boldsymbol{X}\|_{\ell_1} + k\|\boldsymbol{X} - \boldsymbol{A}\|_{\ell_2}^2)$, *then* $x_i^* = \begin{cases} 0, & \text{if } |a_i| \leq \frac{1}{2k} \\ a_i - \frac{sign(a_i)}{2k}, & \text{if } |a_i| > \frac{1}{2k} \end{cases}$

[4] Which essentially is a generalisation of the Lagrange multiplier.

[5] This is usually called shrinkage or soft threshold [163, 13].

The proof of Lemma 3.7.1 starts with the realisation that the vector components x_i can be optimised independently. This involves minimising the real function $f(x_i) = |x_i| + k(x_i - a_i)^2$. Then, the absolute value can be eliminated by noting that x_i and a_i must have the same sign, and then finally a simple quadratic function is minimised.



(a) ℓ_2 (b) ℓ_1

(c) ℓ_2 MAE= 2.2^o (d) ℓ_1 MAE= 1.5^o

Figure 3.6 Quantitative comparison of the ℓ_2 vs ℓ_1 solution for a simple synthetic data experiment (Lambertian reflection/6 images). The top row shows the obtained surfaces using these two solver variations respectively. These surfaces then were numerically differentiated (Equation 3.8) in order to generate normal maps; the difference of the generated normal maps to the ground truth (error in degrees) is shown in the bottom row. Although not very apparent on the 3D reconstructions in (a),(b) (the ℓ_2 is marginally more flattened), the ℓ_1 solver achieves a better quantitative result due to the presence of depth discontinuities which are outliers to the differential model.

A quantitative comparison of the ℓ_1 versus ℓ_2 optimisation schemes is shown in Figure 3.6 which shows a simple synthetic PS dataset on the "Armadillo" model from The Stanford 3D Scanning Repository[6]. This synthetic model has depth discontinuities mostly around the face which violate the normal/depth relation and are outliers for the differential model. Although not very apparent in the 3D reconstruction (Figure 3.6a-3.6b), the normal maps in

[6]http://graphics.stanford.edu/data/3Dscanrep

Figure 3.6c-3.6d reveal a lower error for the ℓ_1 optimisation thus justifying the usage of this solver.

Computational Complexity of ADMM As the main aim of this thesis is practicality, it is worth commenting on the computational complexity of ADMM in contrast to simple least squares solution. Assuming that there are N equations (typically 32 from all the pairs of 8 images) and p pixels in the domain (typically $10^5 - 10^6$), then the system matrix of LST is sparse $p \times p$ with around $9p$ entries (see Equation 3.34) for $G_u^\intercal G_u$ giving $4p$ entries and then $G_v^\intercal G_v$ another $4p$ and then p from the regulariser. It is noted that the matrix to be inverted is constant for all iterations ($A_s^\intercal A_s + \frac{\lambda}{\alpha} I_{p\times p}$) and thus Cholesky factorisation can be used and the result can be cached. Hence, only the first iteration is computationally expensive with subsequent $\boldsymbol{Z}$ updates being orders of magnitude computationally less expensive. Nonetheless, if the number of images is large, the residual update step in Equation 3.40 will be computationally expensive due to the matrix vector multiplication.

More Robust Loss Function As seen in Section 2.2.1, there are a number of more robust ways of dealing with the PS problem than the simple ℓ_1, with different loss functions having different properties. However, the convergence of the ADMM scheme is only guaranteed if the loss function ($\boldsymbol{G}$ in Equation 3.37) is convex (Boyd and Vandenberghe [19]), thus a Cauchy loss function $\left(f(x) = \lambda^2 \log(1 + \frac{x^2}{\lambda^2})\right)$ can not be applied. In addition, the ℓ_1 loss has the simple shrinkage update step in Equation 3.38 which needs to be replaced with a more complicated minimisation in the case of another loss function.

3.7.5 Approximating the shininess parameter

Since the unifying irradiance equation allows a simultaneous parameterization of diffuse and specular reflectance, the approximation of the shininess coefficient c is an important achievement for the material understanding of the depicted scene. Simultaneously directly estimating the linked z and c variables is very challenging, thus an alternating procedure is presented instead. Assuming known z, $\boldsymbol{N}$ can be trivially calculated using finite differences (Equation 3.29) and $\boldsymbol{W}$ using Equation 3.15, assuming for now that we are in the specular zone. This final assumption is used to overcome indirect dependence of $\boldsymbol{W}$ to c. Thus c can be estimated as follows:

Starting again from the ratio of irradiance equations for images j and k (Equation 3.17) these are re-arranged to get:

$$\left(\frac{i_j a_k}{i_k a_j}\right)^{c_{jk}} = \frac{\overline{\boldsymbol{W}}_j \cdot \boldsymbol{N}}{\overline{\boldsymbol{W}}_k \cdot \boldsymbol{N}} \Rightarrow \tag{3.41}$$

$$c_{jk} = \frac{\log\left(\frac{\overline{\boldsymbol{W}}_j \cdot \boldsymbol{N}}{\overline{\boldsymbol{W}}_k \cdot \boldsymbol{N}}\right)}{\log\left(\frac{i_j a_k}{i_k a_j}\right)} \tag{3.42}$$

where the subscript c_{jk} implies that this is the estimate of c from the pair of images j,k.

We note that there are n versions of Equation 3.42, one for each pair of images. To get an overall estimate, we seek to minimise $\min_c(\|[c_1 - c, \ldots, c_n - c]\|)_{\ell_1}$, thus we consider the median over the whole set of pairs, i.e:

$$c(u, v) = \underset{j,k}{\text{median}}\Big(c_{jk}(u, v)\Big). \tag{3.43}$$

It is noted that Equation 3.43 allows for estimating one c value per pixel and thus it is possible to deal with scenes containing objects of various reflectance without performing any segmentation. However, if a scene is known to contain a single object of uniform reflectance, c can be set to the mean of c to get a more accurate estimate.

Not fully specular point The above c update formula is only valid assuming the specific pixel is in the fully specular zone, so $\boldsymbol{W}$ is independent of c. In order to get the overall optimum c, one has to consider the other two cases and finally select c minimising the total error:

$$\min_c\Big(\sum_{j,k} \| \left(\frac{i_j a_k}{i_k a_j}\right)^c - \frac{\overline{\boldsymbol{W}}_j \cdot \boldsymbol{N}}{\overline{\boldsymbol{W}}_k \cdot \boldsymbol{N}} \|_{\ell_1}\Big), \quad c \in \{(3.43), c_T, 1\}. \tag{3.44}$$

In Equation 3.44, c_T is the c in the transition zone. As $\boldsymbol{W}$ depends on c in the transition zone, the estimation of c_T requires iterating between update c_T and $\boldsymbol{W}$ until convergence or divergence of c_T out of the transition zone.

3.7.6 Near field and specular reflection

As shown in the previous section, the shininess parameter cannot be directly calculated and requires an alternating optimisation scheme. This iterative algorithm becomes an even greater necessity in the case of a perspective camera and/or point light sources and/or non Lambertian reflection, as the fields $\boldsymbol{b} = \boldsymbol{b}(z)$ and $s = s(z)$ depend explicitly on the unknown z. This

Algorithm 1: Calibrated Near-Field Photometric Stereo

Input: Images, calibrated light sources, rough mean distance z_0
Output: Depth map $z(u, v)$, shininess parameter map $c(u, v)$
Initialisation:
$z(u, v) = z_0$, $c(u, v) = 1$;
while $\mid z^{t+1} - z^t \mid > 10^{-4} \times \mid z^t \mid$ **do**
  calculate fields $\boldsymbol{W}_k^t(u, v, z^t)$, $a_k^t(u, v, z^t)$ $\forall k$ (Eq 3.13,3.15) ;
  calculate fields $\boldsymbol{b}_{jk}^t(u, v, z^t)$, $s_{jk}^t(u, v, z^t)$ $\forall (j, k)$ (Eq 3.19);
  eliminate black/saturated pixels and fill holes § 3.5.1 ;
  solve for $z^{t+1}\left(\boldsymbol{b}_{jk}^t(u, v, z^t), s_{jk}^t(u, v, z^t)\right)$ (Eq 3.37) ;
  calculate $\boldsymbol{N}^{t+1}(u, v, z^{t+1})$ (Eq 3.29) ;
  calculate fields $\boldsymbol{W}_k^{t+1}(u, v, z^{t+1})$, $a_k^{t+1}(u, v, z^{t+1})$ $\forall k$ (Eq 3.13,3.15) ;
  calculate $c^{t+1}\left(z^{t+1}, a^{t+1}, \mathbf{W}^{t+1}\right)$ § 3.7.5 ;
end

means that the unified PDE 3.18 is quasi-linear (i.e. linear only in terms of the derivative) and thus the discretisation into a simple linear problem is not possible.

To handle this issue, an iterative procedure is followed. This involves calculating a, $\boldsymbol{L}$ and thus $\boldsymbol{b}$, s using the previous estimate of z and then following the procedure of Section 3.7.4 to update z. Note that the overall optimisation is non-convex and that there is no theoretical guarantee of convergence towards a local minimum. However, the rest of the thesis provides experimental evidence that in practice, this procedure quickly converges towards a reasonable solution. The alternating procedure is summarised in Algorithm 1, with the superscript t denoting the iteration number.

3.7.7 Numerically integrating normal maps

The chapter is concluded by presenting a very useful property of the variational solver introduced above. The variational solver can be used to robustly integrate normal maps, which will prove to be essential in Chapters 4 and 5 for comparing the proposed methods, which output depth maps directly, to other state-of-the-art approaches which generate normal maps.

Starting from the classic PS normal parameterisation:

$$\left[\boldsymbol{N}^1, \boldsymbol{N}^2, \boldsymbol{N}^3\right]^\intercal \propto \left[\frac{\partial z(u,v)}{\partial u}, \frac{\partial z(u,v)}{\partial v}, -1\right]^\intercal, \tag{3.45}$$

$$\left[-\frac{\boldsymbol{N}^1}{\boldsymbol{N}^3}, -\frac{\boldsymbol{N}^2}{\boldsymbol{N}^3}, -1\right]^\intercal = \left[\frac{\partial z(u,v)}{\partial u}, \frac{\partial z(u,v)}{\partial v}, -1\right]^\intercal, \tag{3.46}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \nabla z = - \begin{bmatrix} \frac{\boldsymbol{N}^1}{\boldsymbol{N}^3} \\ \frac{\boldsymbol{N}^2}{\boldsymbol{N}^3} \end{bmatrix}. \tag{3.47}$$

Equation 3.47 is of the same form as the PS fundamental Equation 3.18 and thus can be solved robustly on any image domain with the variation optimisation presented in Sections 3.7.2 to 3.7.4. This is better than the classic normal integration methods described in Section 2.2.2 (Frankot and Chellappa [48], Simchony et al. [168]). Note that if perspective projection is to be used, the above parameterisation breaks down as the normal parametrisation depends on the unknown depth. One has to use the substitution $g = \log(z + f)$, as described in Section 3.3.1, or alternatively, use an iterative scheme like Algorithm 1.

Chapter 4

Photometric Stereo in Ambient Light

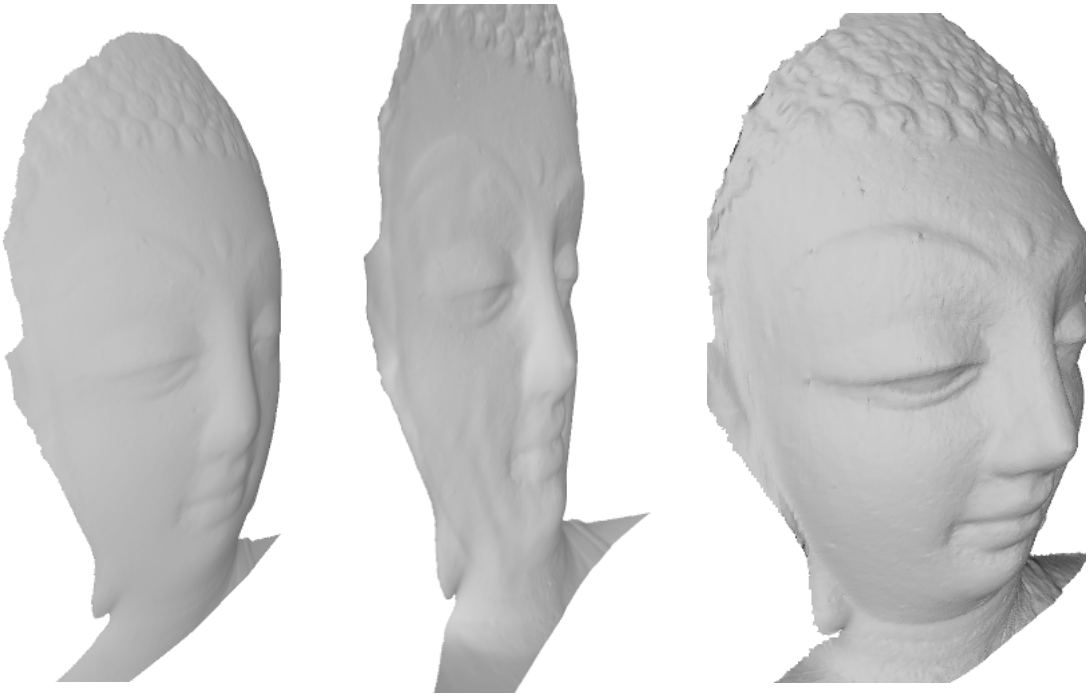
Figure 4.1 The problem of Photometric Stereo in Ambient Light: (*left*) ambient light image (i.e. no PS light sources on) which significantly reduces the photometric parallax for images taken with different illumination (*middle*, *right* are 2 examples out of a sequence of 24 images).

Most of the literature of PS assumes that the total amount of light captured by the camera is due to the reflection through the object, modelled through the image irradiance equation (see Section 2.2). Generally, this is a very reasonable assumption, especially when the acquisition takes place in a controlled lab setting. However, as the attention shifts into the wild, the need for dealing with non-negligible *ambient light* emerges. The concept of ambient light basically covers all the unexplained secondary reflections between the surface and the environment, which may either be due to stray lights, or inter-reflections or even back scattering, and are thus difficult or impossible to model in closed-form. Such deviations induce a systematic additive bias of the linear model which can be modelled as (Equation 4.1) where $e(u, v)$ represents the unexplained bias that could be dependent on the light source $\boldsymbol{L}$ itself as:

$$i(u, v) = i_{\mathbf{BRDF}}(\mathbf{L}, \mathbf{N}, \mathbf{V}) + e(u, v). \tag{4.1}$$

In this chapter, a PS method which deals with ambient light is introduced. This method keeps all the realistic modelling assumptions introduced in Chapter 3; these include perspective projection, nearby point light sources and specular reflection.

4.1 Motivation



(a) Ikehata et al. [72] (b) Mecca et al. [118] (c) Proposed

Figure 4.2 Illustration of the flattening effect that ambient light causes on state-of-the art methods applied on the dataset of Figure 4.1.Ikehata et al. [72] employed a very robust sparse Bayesian regression technique that is completely unable to cope with the systematic bias caused by the ambient lighting. (b) illustrates the baseline framework method presented in Chapter 3 and again its result is severely deformed by the ambient light bias. On the contrary, (c) shows that the novel formulation proposed in this chapter is able to cope with this issue.

The motivation for choosing to tackle ambient light was prompted by two recent applications of PS: real time 3D reconstruction by Hernández et al. [58], and PS in open environments (outdoor PS Hung et al. [70], Abrams et al. [1], as well as PS in a scattering medium such as murky water Murez et al. [126]). The usual approach taken to deal with ambient light consists of capturing an image without any artificial lights first, and then substracting it from all the images. However, it is not possible to apply such preprocessing in real time PS, where the shape information is extracted from a single RGB image of a

commercial camera sensor, Hernández et al. [58], Brostow et al. [21]. In addition, it is impossible to have a priori knowledge of the ambient light if it only appears when the artificial lights are on, as is the case with back scattering in murky water, Treibitz and Schechner [179], Tsiotsios et al. [181]. To deal with these issues, a framework for PS in ambient light which does not require any preprocessing step is introduced.

The assumed model for ambient light consists of a non-uniform scalar field, which is independent from $\boldsymbol{L}$. Although it cannot be physically justified that mutual reflections, or backscatter lightings, are independent of the lighting, this approximation is numerically tractable, and was already used successfully in [205, 181, 126]. In the literature, such an ambient term was accounted for using two different strategies: offline calibration or online estimation. The first strategy simply involves capturing an image in the "dark" (i.e. with no active lights turned on). This "ambient image" can then be subtracted from the PS images to create "ambient-free" images.

The other strategy comprises estimating $e(u, v)$ along with the surface characteristics. The classic pixelwise PS estimation is simple to extend: the new recovery problem is another linear system of diffuse irradiance equations accounting $e(u, v)$ as unknown, whose solution is uniquely defined as long as the number of images is at least 4, and the rank of the system is 4. More specifically, this re-writes Equation 2.13 as:

$$[L_s, \mathbf{1}] \begin{bmatrix} \boldsymbol{M}(u, v) \\ e(u, v) \end{bmatrix} = \boldsymbol{I}_s(u, v). \tag{4.2}$$

In Equation 4.2, the fourth component of the unknown vector is the ambient light, and its first three components represent the surface normals scaled by the albedo. This is essentially the same strategy as the approach of Yuille et al. [205], although they extended this procedure to the *uncalibrated* PS problem (see Section 5.2.1). However, this approach relies on the restrictive assumptions of directional lighting as well as a Lambertian surface reflection, and it performs poorly if these assumptions are violated (see comparison with [205] in Section 4.3).

Another relatively simple way of accounting for the ambient light bias is to use a spherical harmonics model for the lighting as proposed by Basri and Jacobs [11]. Such a model was used by Basri et al. [12] to tackle the problem of unknown and non-directional lighting for a Lambertian object. Using spherical harmonics decomposition, it was shown that the surface normals lie on a low-dimensional space spanned by the principal components of the image. In particular, a four harmonics model may describe the combination of directional lighting and "ambient" lighting rather well. Such a model was recently used by Or-el et al. [139] in order to apply SFS under "natural" illumination. Yet, such spherical harmonics models prevent one from using closed form expressions for non-directional lightings, e.g. the pointwise

source model accounting for radial propagation of light and inverse of squared distance light attenuation.

4.2 Dynamic ambient light removal

To start, recall the irradiance Equation 3.16 introduced in Chapter 3 that models perspective viewing geometry, radial propagation, light attenuation and both diffuse and specular reflection and add an additional pixelwise unknown $e(x, y)$, the ambient light. Thus the overall irradiance equation for the j^{th} light source is:

$$i_j(u, v) = \rho(u, v)a_j(u, v, z(u, v))(\overline{\boldsymbol{N}}(u, v) \cdot \overline{\boldsymbol{W}}_j(u, v, z(u, v)))^{\frac{1}{c(u,v)}} + e(u, v). \tag{4.3}$$

The terms $\rho(u, v)$, $a_j(u, v, z(u, v))$, $\boldsymbol{W}_j(u, v, z(u, v))$ and $c(u, v)$ follow the same definitions introduced in Section 3.5. The irradiance Equation 4.3 now consists of two terms, and thus the ratio method presented in Section 3.2 cannot be used in order to cancel out the albedo and the non-linear denominator of the normal. In addition, note that if $c = 1$, i.e. the object is purely Lambertian, it is possible to consider differences of images i.e. $i_j - i_k$ for images j, k and then follow the ratio procedure, thus simultaneously cancelling both the ambient light term as well as the albedo and non-linear denominator of the normal. However, in the most general setting, $c \leq 1$, and thus the difference of the two exponentials is hard to handle analytically. This requires special algebraic manipulations as explained below.

With the aim being the consideration of readable equations, the notation is simplified, and as far as possible, writing the dependencies of the functions is avoided from now on.

First of all, the irradiance equation is re-arranged as follows:

$$(i_j - e)^c = (\rho a_j)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j. \tag{4.4}$$

Considering the binomial expansion of the left hand side:

$$(i_j - e)^c = i_j^c - cei_j^{c-1} + \frac{c(c-1)}{2}e^2 i_j^{c-2} + \dots \tag{4.5}$$

Truncating this expansion to the first order, and substituting to Equation 4.4, the following holds:

$$i_j^c - cei_j^{c-1} \approx (\rho a_j)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j \tag{4.6}$$

$$\Downarrow$$

$$i_j - ce \approx \rho^c \gamma_j \frac{\boldsymbol{N}}{|\boldsymbol{N}|} \cdot \overline{\boldsymbol{W}}_j \tag{4.7}$$

where the alias $\gamma_j = \frac{a_j^c}{i_j^{c-1}}$ is used.

It is noted that in Equation 4.7, in the pure diffuse case (i.e. $c = 1$), there is no need for binomial expansion and truncation and thus Equation 4.7 is actually an exact equality. This makes the model more accurate in the framework upon which most of the research about PS focuses. In addition, as $i_j > e$, the binomial expansion is always convergent and thus the truncation operation is mathematically valid. The accuracy of the truncation is further examined in Section 4.2.3.

Now, Equation 4.7 is applied to two pairs of irradiance equations, namely the j^{th}, k^{th} and the q^{th}, r^{th}. Considering the ratio of their differences and this operation lets the ambient light cancel out together with the albedo and the normalisation of the normal as:

$$\frac{i_j - ce - i_k + ce}{i_q - ce - i_r + ce} \approx \frac{\frac{\rho^c}{|\boldsymbol{N}|}\left[\gamma_j \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_j - \gamma_k \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_k\right]}{\frac{\rho^c}{|\boldsymbol{N}|}\left[\gamma_q \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_q - \gamma_r \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_r\right]} \tag{4.8}$$

that is

$$\frac{i_{jk}}{i_{qr}} = \frac{i_j - i_k}{i_q - i_r} \approx \frac{\gamma_j \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_j - \gamma_k \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_k}{\gamma_q \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_q - \gamma_r \boldsymbol{N} \cdot \overline{\boldsymbol{W}}_r} \tag{4.9}$$

where the aliases $i_{jk} = i_j - i_k$ and $i_{qr} = i_q - i_r$ are used.

By substituting the differential parameterization of the normal from Equation 3.8, Equation 4.9 yields the following quasi-linear PDE:

$$\boldsymbol{b}_u(u, v, z(u, v))\frac{\partial z}{\partial u} + \boldsymbol{b}_v(u, v, z(u, v))\frac{\partial z}{\partial v} = s(u, v, z(u, v)) \tag{4.10}$$

with:

$$\boldsymbol{b}_u = i_{qr}\left[\gamma_j\left(\overline{\boldsymbol{W}}^1_j - \frac{u}{f}\overline{\boldsymbol{W}}^3_j\right) - \gamma_k\left(\overline{\boldsymbol{W}}^1_k - \frac{u}{f}\overline{\boldsymbol{W}}^3_k\right)\right] - i_{jk}\left[\gamma_q\left(\overline{\boldsymbol{W}}^1_q - \frac{u}{f}\overline{\boldsymbol{W}}^3_q\right) - \gamma_r\left(\overline{\boldsymbol{W}}^1_r - \frac{u}{f}\overline{\boldsymbol{W}}^3_r\right)\right], \tag{4.11}$$

$$\boldsymbol{b}_v = i_{qr}\left[\gamma_j\left(\overline{\boldsymbol{W}}^2_j - \frac{v}{f}\overline{\boldsymbol{W}}^3_j\right) - \gamma_k\left(\overline{\boldsymbol{W}}^2_k - \frac{v}{f}\overline{\boldsymbol{W}}^3_k\right)\right] - i_{jk}\left[\gamma_q\left(\overline{\boldsymbol{W}}^2_q - \frac{v}{f}\overline{\boldsymbol{W}}^3_q\right) - \gamma_r\left(\overline{\boldsymbol{W}}^2_r - \frac{v}{f}\overline{\boldsymbol{W}}^3_r\right)\right], \tag{4.12}$$

$$s = \frac{f+z}{f}\left[i_{qr}\left(\gamma_j\overline{\boldsymbol{W}}^3_j - \gamma_k\overline{\boldsymbol{W}}^3_k\right) - i_{jk}\left(\gamma_q\overline{\boldsymbol{W}}^3_q - \gamma_r\overline{\boldsymbol{W}}^3_r\right)\right]. \tag{4.13}$$

It is noted that Equation 4.10 has exactly the same form as the fundamental PS Equation 3.18 and thus can be solved using the variational optimisation framework described in Section 3.7.

Specularity parameter estimation In order to update the specularity parameter c, the ratio of differences of the irradiance Equation 4.3 for images j, k to q, r, is considered in order to get:

$$\frac{i_j - i_k}{i_q - i_r} = \frac{a_j(\boldsymbol{N}\cdot\overline{\boldsymbol{W}}_j)^{\frac{1}{c}} - a_k(\boldsymbol{N}\cdot\overline{\boldsymbol{W}}_k)^{\frac{1}{c}}}{a_q(\boldsymbol{N}\cdot\overline{\boldsymbol{W}}_q)^{\frac{1}{c}} - a_r(\boldsymbol{N}\cdot\overline{\boldsymbol{W}}_r)^{\frac{1}{c}}}. \tag{4.14}$$

Expanding out Equation (4.14), gives:

$$(i_j - i_k)a_q(\overline{\boldsymbol{N}}\cdot\overline{\boldsymbol{W}}_j)^{\frac{1}{c}} - (i_j - i_k)a_r(\overline{\boldsymbol{N}}\cdot\overline{\boldsymbol{W}}_k)^{\frac{1}{c}} - (i_r - i_q)a_j(\overline{\boldsymbol{N}}\cdot\overline{\boldsymbol{W}}_q)^{\frac{1}{c}} + (i_r - i_q)a_k(\overline{\boldsymbol{N}}\cdot\overline{\boldsymbol{W}}_r)^{\frac{1}{c}} = 0 \tag{4.15}$$

which is of the form $\sum_{n=1}^{4} d_n m_n^x = 0$, with d_n, m_n known scalars and $x = \frac{1}{c}$. Equation 4.15 cannot be solved analytically; it can, however, be solved numerically with the Newton-Raphson method.

4.2.1 Parameterisation limitations

After having established a new differential formulation for the PS problem in ambient light, this section presents some cases where the ratio of differences parameterisation of Equations 4.11 to 4.13 breaks down or becomes numerically unstable. In general, for a well-posed problem, $\|\boldsymbol{b}\| > 0$, and the higher this magnitude is, the better the numerical stability. In addition, $\boldsymbol{b}$ from different quadruples of images have to be linearly independent (these properties are true for the standard parameterisation presented in Chapter 3 as guaranteed by Theorems 3.6.1 and 3.6.2).

It is important to note that moving the assumptions away from the classic PS yields a more non-linear problem as $\boldsymbol{b}$ becomes more chaotic and hence less likely to be ill-constrained. This is not surprising, as it is a well known fact that the near-field non-linearities remove ambiguities. The most characteristic example for that is Papadhimitri and Favaro [142], where it was proven that the near-field un-calibrated PS does not suffer from the Bass-Relief ambiguity. In addition, even collinear light sources can lead to a well-posed problem assuming the γ factor (see Equation 4.7) has enough change between views.

Nonetheless, assuming the classic PS assumptions are true, i.e. $f = \infty$, $c = 1$, $a = 1$ and $\boldsymbol{L}$ constant for all pixels (i.e. directional lighting), then Equations 4.11 to 4.13 reduce to:

$$\boldsymbol{b} = \begin{bmatrix} i_{qr}(\overline{\boldsymbol{L}}_j^1 - \overline{\boldsymbol{L}}_k^1) - i_{jk}(\overline{\boldsymbol{L}}_q^1 - \overline{\boldsymbol{L}}_r^1) \\ i_{qr}(\overline{\boldsymbol{L}}_j^2 - \overline{\boldsymbol{L}}_k^2) - i_{jk}(\overline{\boldsymbol{L}}_q^2 - \overline{\boldsymbol{L}}_r^2) \end{bmatrix} \quad \text{and} \quad s = i_{qr}(\overline{\boldsymbol{L}}_j^3 - \overline{\boldsymbol{L}}_k^3) - i_{jk}(\overline{\boldsymbol{L}}_q^3 - \overline{\boldsymbol{L}}_r^3). \tag{4.16}$$

Equation 4.16 reveals another corner case: if all the light sources $\overline{\boldsymbol{L}}_i^3 = \overline{\boldsymbol{L}}_j^3 = \overline{\boldsymbol{L}}_q^3 = \overline{\boldsymbol{L}}_r^3$ then $s = 0$ and thus Equation 4.16 only describes the level set of the surface. In fact, a similar property has been used by Alldrin and Kriegman [5] and Zhou et al. [209] in order to start from a sparse point cloud and propagate along the depth iso-contours to create a dense surface, in a multi-view setting (see Chapter 6). This limitation is actually relevant, as the experiment setup introduced in Figure 1.5 contains all the light sources on plane parallel to the image plane. Thus, to get around this problem, the LEDs are placed in circles with significantly different radii. This means that the distance between the light sources and the scene is significantly different for LEDs from different circles and thus the light attenuation a (and thus subsequently γ, are different) so the simplification leading to Equation 4.16 does not hold. Finally, a semi-obvious fact is noted, that, as the model is generated using a ratio of differences, it becomes invalid in the case of equality of images $j = k$ or $q = r$.

4.2.2 Second order expansion

In this section, it is shown how more than four images can be used to cancel up to the second order remainder of the binomial expansion and increase the accuracy of the truncation.

Start by taking into account the expansion till the second order:

$$i_j^c - c\, e\, i_j^{c-1} + \frac{c(c-1)}{2}\, e^2\, i_j^{c-2} \approx (\rho a_j)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j \tag{4.17}$$

and combining the previous equation for images j^{th} and k^{th} as $\frac{(4.17)_j}{i_j^{c-1}} - \frac{(4.17)_j}{i_k^{c-1}}$ to give:

$$\begin{aligned}\frac{i_j^c - c\,e\,i_j^{c-1} + \frac{c(c-1)}{2}\,e^2\,i_j^{c-2}}{i_j^{c-1}} - \frac{i_k^c - c\,e\,i_k^{c-1} + \frac{c(c-1)}{2}\,e^2\,i_j^{c-2}}{i_k^{c-1}} \approx \\ \frac{(\rho a_j)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j}{i_j^{c-1}} - \frac{(\rho a_k)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_k}{i_k^{c-1}}\end{aligned} \tag{4.18}$$

that simplifies to:

$$i_j + \frac{c(c-1)}{2}\frac{e^2}{i_j} - i_k - \frac{c(c-1)}{2}\frac{e^2}{i_k} \approx \rho^c \gamma_j \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j - \rho^c \gamma_k \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_k \tag{4.19}$$

that leads to:

$$i_j - i_k + \frac{c(c-1)}{2}\,e^2\,\left(\frac{1}{i_j} - \frac{1}{i_k}\right) \approx \rho^c \left(\gamma_j \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j - \gamma_k \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_k\right) \tag{4.20}$$

where again the alias $\gamma_j = \frac{a_j^c}{i_j^{c-1}}$ and the equivalent for the other indexes, is used. In addition, the additional auxiliary variables $i_{jk} = i_j - i_k$ and $\boldsymbol{H}_{jk} = \gamma_j \overline{\boldsymbol{W}}_j - \gamma_k \overline{\boldsymbol{W}}_k$ are defined. Then multiply both sides of Equation 4.20 by the factor $\frac{i_j\,i_k}{i_j - i_k}$ to get:

$$i_j\,i_k - \frac{c(c-1)}{2}e^2 \approx \rho^c\,\frac{i_j\,i_k}{i_{jk}}\,\overline{\boldsymbol{N}} \cdot \boldsymbol{H}_{jk}. \tag{4.21}$$

Finally, the following ratio of differences considering the images m^{th}, j^{th}, k^{th} and p^{th}, q^{th}, r^{th} is performed as $\frac{(4.21)_{mj} - (4.21)_{mk}}{(4.21)_{pq} - (4.21)_{pr}}$ to get:

$$\frac{\left(i_m\,i_j - \frac{c(c-1)}{2}\,e^2\right) - \left(i_m\,i_k - \frac{c(c-1)}{2}\,e^2\right)}{\left(i_p\,i_q - \frac{c(c-1)}{2}\,e^2\right) - \left(i_p\,i_r - \frac{c(c-1)}{2}\,e^2\right)} \approx \frac{\left(\rho^c\,\frac{i_m\,i_j}{i_{mj}}\overline{\boldsymbol{N}} \cdot \boldsymbol{H}_{mj}\right) - \left(\rho^c\,\frac{i_m\,i_k}{i_{mk}}\overline{\boldsymbol{N}} \cdot \boldsymbol{H}_{mk}\right)}{\left(\rho^c\,\frac{i_p i_q}{i_{pq}}\overline{\boldsymbol{N}} \cdot \boldsymbol{H}_{pq}\right) - \left(\rho^c\,\frac{i_p\,i_r}{i_{pr}}\overline{\boldsymbol{N}} \cdot \boldsymbol{H}_{pr}\right)} \tag{4.22}$$

that is

$$\frac{i_m\,i_j - i_m\,i_k}{i_p\,i_q - i_p\,i_r} \approx \frac{\frac{i_m\,i_j}{i_{mj}}\boldsymbol{N} \cdot \boldsymbol{H}_{mj} - \frac{i_m\,i_k}{i_{mk}}\boldsymbol{N} \cdot \boldsymbol{H}_{mk}}{\frac{i_p\,i_q}{i_{pq}}\boldsymbol{N} \cdot \boldsymbol{H}_{pq} - \frac{i_p\,i_r}{i_{pr}}\boldsymbol{N} \cdot \boldsymbol{H}_{pr}} \tag{4.23}$$

which simplifies by cancelling the i_m/i_p term to:

$$\frac{i_{jk}}{i_{qr}} = \frac{i_j - i_k}{i_q - i_r} \approx \frac{\frac{i_j}{i_{mj}}\boldsymbol{N} \cdot \boldsymbol{H}_{mj} - \frac{i_k}{i_{mk}}\boldsymbol{N} \cdot \boldsymbol{H}_{mk}}{\frac{i_q}{i_{pq}}\boldsymbol{N} \cdot \boldsymbol{H}_{pq} - \frac{i_r}{i_{pr}}\boldsymbol{N} \cdot \boldsymbol{H}_{pr}} \tag{4.24}$$

$$\frac{i_{mj}\, i_{mk}\, i_{jk}}{i_{pq}\, i_{pr}\, i_{qr}} \approx \frac{i_j\, i_{mk} \boldsymbol{N} \cdot \boldsymbol{H}_{mj} - i_k\, i_{mj} \boldsymbol{N} \cdot \boldsymbol{H}_{mk}}{i_q\, i_{pr} \boldsymbol{N} \cdot \boldsymbol{H}_{pq} - i_r\, i_{pq} \boldsymbol{N} \cdot \boldsymbol{H}_{pr}} \tag{4.25}$$

Again, expanding Equation 4.25 and substituting the differential parameterisation of the normal from Equation 3.8 gives the standard quasilinear PDE form.

4.2.3 Binomial series truncation error

The algebraic manipulation described in Section 4.2 assumes that the first or second order truncation of the binomial series of Equation 4.5 is a fairly good approximation of the actual series. In this section, a more in depth investigation of the truncation error is performed.

More precisely, the error metric is the absolute difference between the infinite series and its first or second order truncation. In addition, the error is more meaningful if expressed as a ratio (expressed as a percentage) with respect to the series value. Thus, ignoring the subscript j, we define the following errors:

$$Err_1 = \frac{(i-e)^c - \left(i^c - c\, e\, i^{c-1}\right)}{(i-e)^c} \tag{4.26}$$

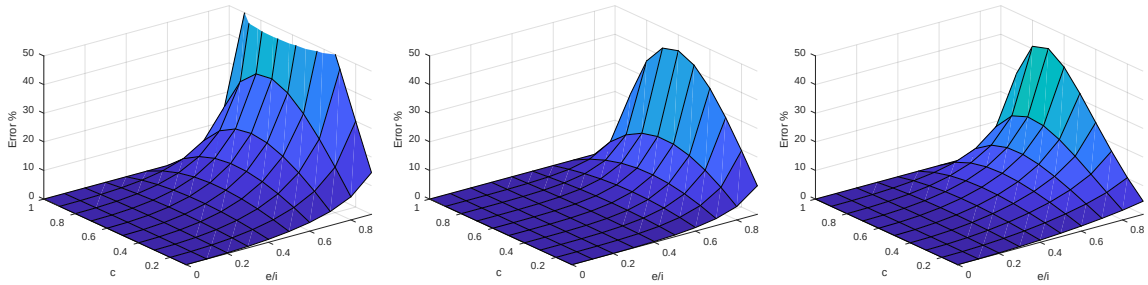
$$Err_2 = \frac{(i-e)^c - \left(i^c - c\, e\, i^{c-1} + \frac{c(c-1)}{2}\, e^2\, i^{c-2}\right)}{(i-e)^c} \tag{4.27}$$

which are re-arranged so that it is clear that they are functions of c and e/i only:

$$Err_1 = \frac{(1-e/i)^c - \left(1 - c\, e/i\right)}{(1-e/i)^c} \tag{4.28}$$

$$Err_2 = \frac{(1-e/i)^c - \left(1 - c\, e/i + \frac{c(c-1)}{2}\, (e/i)^2\right)}{(1-e/i)^c}. \tag{4.29}$$

The magnitude of the truncation errors is illustrated in Figure 4.3. It is clear that unless e/i is unreasonably large (which would mean that the signal-to-noise ratio is too small), the two truncation errors are only a few percentage points. As a numerical example, for $c = 0.5$ and $e/i = 0.25$ (corresponding to signal-to-noise ratio of $3:1$), we have that $Err_1 = 1.04\%$ and $Err_2 = 0.13\%$. Finally, Figure 4.3c shows the difference between the first and second order (i.e. the second order term) and motivates for using the first order formula only. There is marginal theoretical benefit for using the second order term for a significant increase of the computational complexity and the associated numerical errors. It is important to stress that



(a) Order 1 error. (b) Order 2 error. (c) Order 1 error − order 2 error.

Figure 4.3 Illustration of percentage error induced due to the truncation of the binomial series. (a) is the first order error defined in 4.28, (b) is the second order error defined in 4.29 and (c) is their absolute difference $|Err_1 - Err_2|$. It is noted that as most image data are at most 10 bits of precision, any error smaller than 0.1% is computationally zero.

as most image data are at most 10 bits of precision, any error smaller than 0.1% is negligible. Hence the absolute difference between Err_1 and Err_2 is more meaningful than their ratio.

4.3 Experiments

The proposed algorithm is evaluated on both synthetic and real data. Two different capture setups were used for capturing real data. The first setup consisted of a previous version of the plane of LEDs (Figure 1.5); this version had a FL3-U3-20E4C camera placed in the center of a 5cm base containing 24 LEDs (SHARP-MINIZENIGATA). The LEDs were arranged in 2 concentric rings of radii 3cm and 5cm respectively with 8LEDs in the inner circle and 16 in the outer. It must be emphasised that the numerical stability of the proposed approach critically depends on not having all LEDs in a circle around the camera as discussed in Section 4.2.1. This setup was used to capture the marble Buddha statue (Figure 4.1) and a shiny plastic head (Figure 4.4a) datasets and the image resolution was set to 800×600.

The second setup, consisting of a series of 8 LEDs and a Canon EOS 7D camera, was used to acquire images of a plaster Arlequin mask (Figure 4.4b) and of a plaster print of teeth (Figure 4.4c). For both setups, the camera intrinsics were calibrated using the camera calibration toolbox for Matlab [18], the position of the LEDs was calibrated using a mirror sphere, and their intensities were calibrated using a diffuse sphere. The anisotropy parameter was set to $\mu = 1$, which was consistent with the manufacturer's specification.

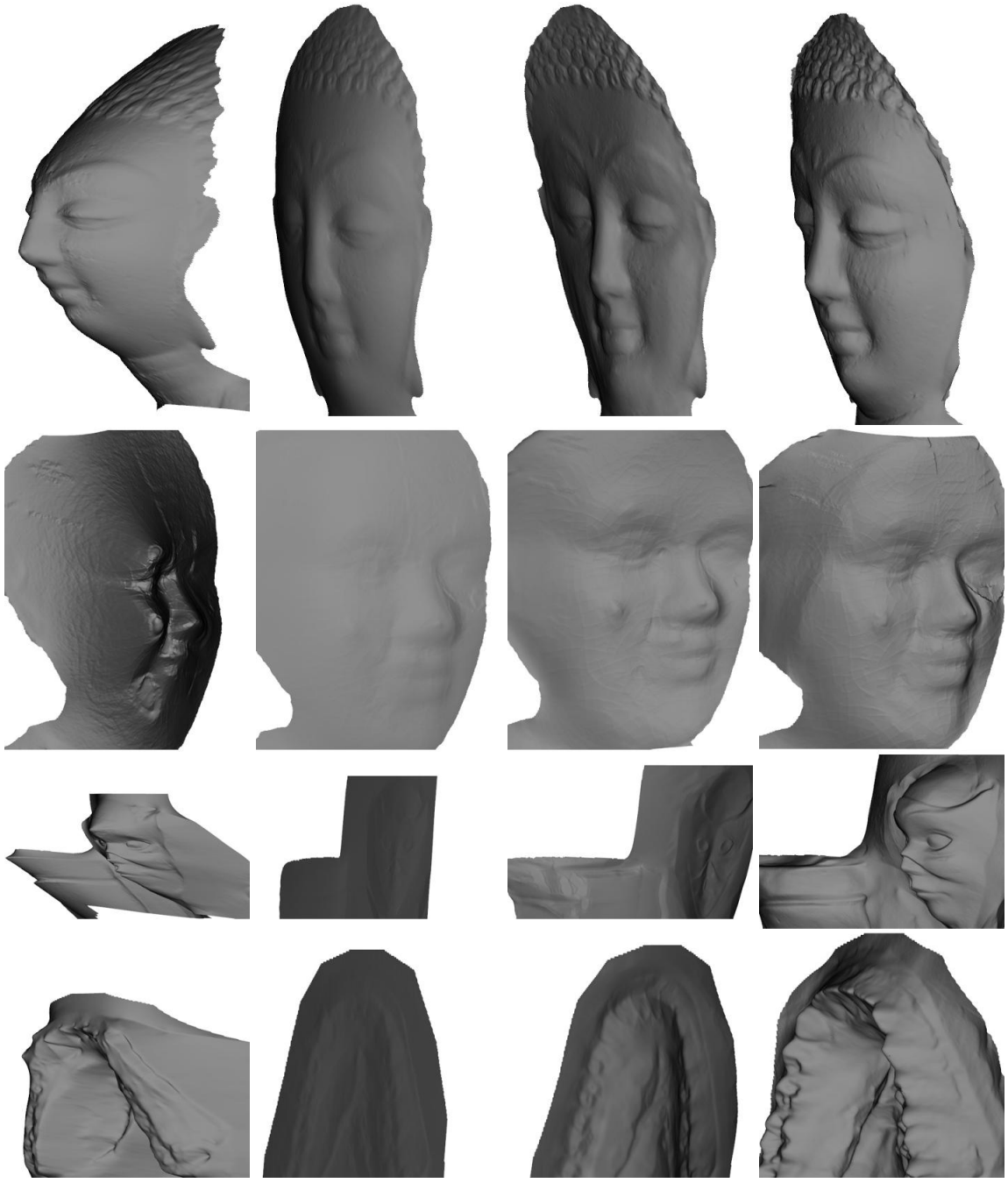
To create challenging conditions, the relevant amount of ambient light was varied between the different datasets, with the Buddha being illuminated more than the Mask and the Teeth. Finally, the shiny head was placed next to an open window which makes the ambient light

(a) Shiny head (b) Mask (c) Teeth

Figure 4.4 Real data with significant ambient Light. These are a shiny plastic head (a), a plaster Arlequin mask (b) and a plaster print of Teeth (c). Two samples from each object and the respective ambient light in the last row. The ambient light shown here is only for demonstration purposes and is not used in any of the computations.

uncontrolled (the reflection of the window is clearly visible and it creates some artifacts in the reconstructions in the second row of Figure 4.5).

Implementation The algorithm was implemented in Matlab and was tested on an i7 processor at 2.4GHz, with 16GB RAM. For most experiments the Order 1 version of the proposed method was used (Equation 4.10). A comparison of Order 1 versus Order 2 is also performed later in Table 4.1. In addition, as selecting all possible quadruples of images is computationally intractable ($\binom{24}{4} = 10626$)), for the first dataset only differences of diametrically opposite images were considered. This maximizes photometric parallax and reduces the number of quadruples to 32 (8 pairs from the outer ring getting divided by 4 pairs from the inner ring). The CPU time on these datasets was a few minutes and the memory usage around 5GB. Finally, the result of the Order 2 method on the Buddha dataset can be found in Figure 7.1.



(a) [205] (b) [72] (c) [118] (d) Proposed

Figure 4.5 Qualitative evaluation of PS methods on the datasets of Figures 4.1 and 4.4. [118] is the baseline theoretical framework described in Chapter 3 that does not model the ambient light and thus the reconstruction is severely flattened. This is also similar to the result of [72] which employs a robust estimation strategy that cannot deal with the systematic bias of the ambient light. Finally, [205] specifically models the ambient light but cannot deal with near field effects and specular reflections and thus the result is deformed as well.

Comparisons The proposed method is compared against Yuille et al. [205], Ikehata et al. [72] and Mecca et al. [118]. From these three methods, only Yuille et al. [205] models the ambient light explicitly; Ikehata et al. [72] is one of the most robust methods for solving the classic PS problem. Finally, a comparison with Mecca et al. [118] (which is the basis of the framework presented in Chapter 3) is provided as a baseline test, demonstrating the importance of explicitly handling the ambient light term. For Ikehata et al. [72], their original code was used whereas for Yuille et al. [205] a simple least squares solution of Equation 4.2 was implemented. In addition, the variational solver (see Section 3.7.7) was used in order to numerically integrate the normal maps produced by [205, 72] and thus be able to compare 3D surfaces.

Experimental results The real data experiments on the datasets of Figure 4.1 and Figure 4.4 are shown in Figure 4.5. The proposed algorithm (Order 1) outperforms all competing methods especially [72, 118] which do not model the ambient light explicitly. This confirms the importance of that particular work. In addition, [205] also suffers from some deformations due to the fact that it does not model near field effects and specular reflections which are essential in a number of realistic situations.

Although challenging, these real data experiments only provide qualitative evaluation due to not having access to reliable ground truth data. Therefore, the method is also evaluated on a large number of synthetic data experiments as explained in the next section.

4.3.1 Synthetic data

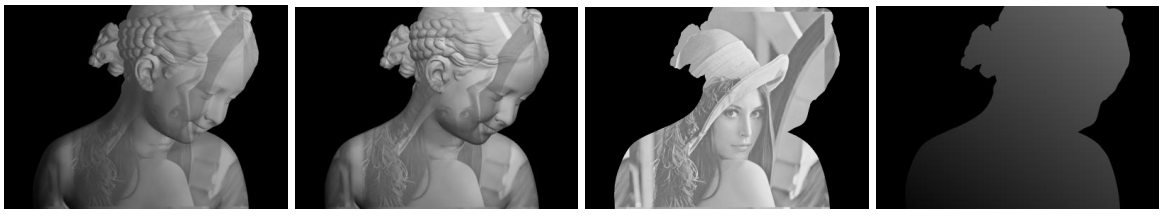
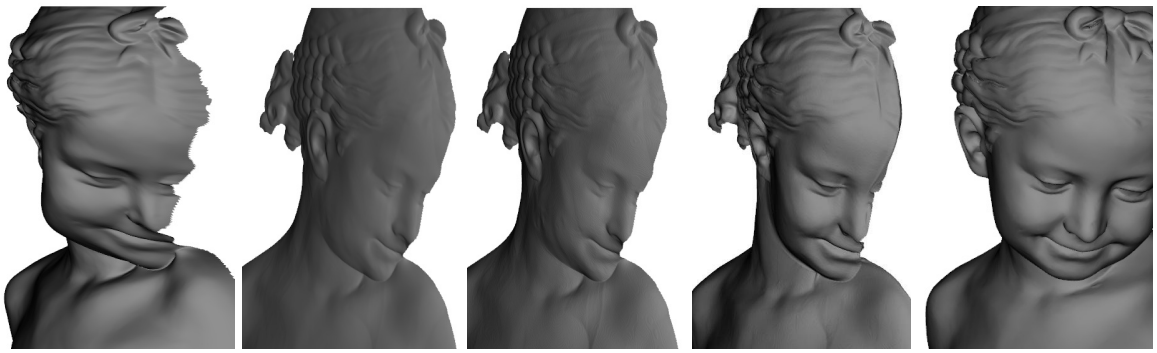


Figure 4.6 Baseline synthetic data: 2 out of the 24 images of our simplified synthetic dataset rendered with orthographic geometry, directional lighting and Lambertian reflection. This dataset offers a fair comparison with [205, 72]. The third image shows the albedo and the fourth one the ambient light.

In order to quantitatively evaluate the method in a number of different situations synthetic data experiments are considered. The model used as a synthetic object was “Bimba” from the AIM@Shape Repository. All datasets (expect Figure 4.10) used 24 images of resolution

800x600x24bits[1] in a configuration closely resembling that of the first real data setup. In addition, the Order 1 version of the method is used considering the same pairs of images as for real data.

Evaluation metric Quantitative evaluation is performed by comparing the generated normal maps with the ground truth (Figure 4.8). The evaluation metric is *MAE*, the mean angular error. Most significantly, it is important to note that comparison is far more meaningful on the normal maps than on the depth maps or 3D meshes which is the standard for other 3D reconstruction methods (e.g. Scharstein and Szeliski [160] and Seitz et al. [162]). This is because the photometric information is encoded in normals and any numerical integration process, either classical (Section 2.2.2) or variational (Section 3.7) will inevitably have biases, especially around depth discontinuities (more about that in Section 6.3.1). Therefore, for the proposed method as well as [118] normals need to be calculated using finite differences of the depth map (Equation 3.29 and 3.8). In addition, even though Yuille et al. [205] and Ikehata et al. [72] provide depth maps, these are not-integrable and do not correspond to valid surfaces. Therefore, in order to have a fair comparison, numerical integration is performed again (Section 3.7.7) and then numerical differentiation.



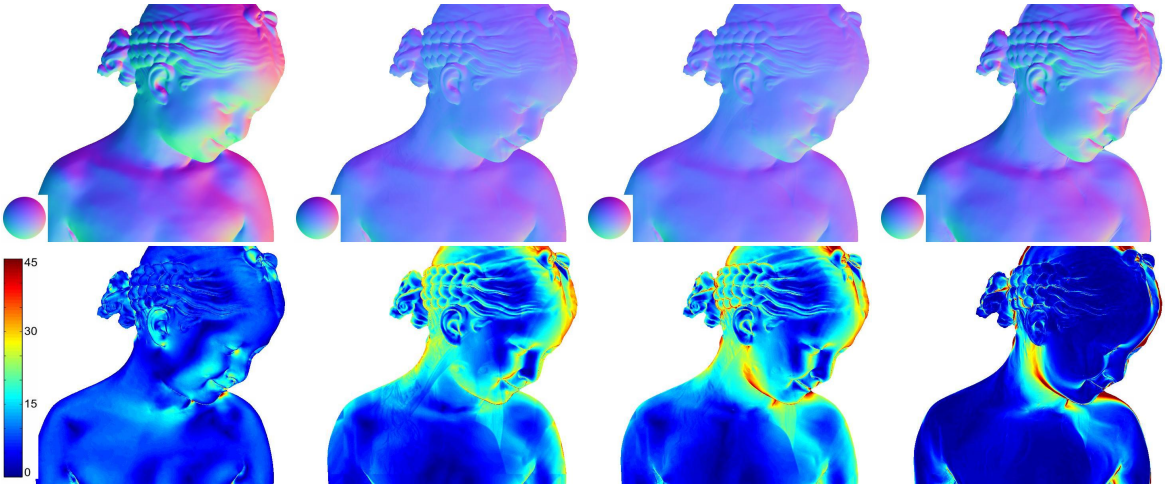
(a) [205] (b) [72] (c) [117] (d) Proposed (e) GT

Figure 4.7 Qualitative comparison of several PS methods on the dataset of Figure 4.6. (e) shows the ground truth. Note that (b) and (c) exhibit a substantial flattening due to the ambient light.

The first synthetic data experiment contained data generated with a simplified image formation model (Figure 4.6) including orthographic viewing geometry, directional illumination and Lambertian reflection. This was done in order to have a fair comparison with Yuille et al. [205] and Ikehata et al. [72]. The ambient light was set to grow linearly, from bottom-left to

[1]Note that most real cameras provide 10bit data, hence these synthetic data are of much higher precision.

top-right, to up to 45% of the maximum intensity value. The "Lena" image was used for the albedo. Figure 4.7 shows the reconstructions obtained, offering qualitative comparison; Figure 4.8 shows the normal maps which provide a quantitative evaluation.



(a) [205],MAE= 11^o (b) [72],MAE= 16.7^o (c) [118],MAE= 17.2^o (d) Ours, MAE= 8.5^o

Figure 4.8 Quantitative evaluation of several PS methods on the dataset of Figure 4.6. Top row shows the resulting normal maps and bottom row shows the map of the error in normals (in degrees) compared to the ground truth. It is noted that most errors are observed around the occlusion boundaries as expected.

The proposed approach clearly outperforms [72, 118] as these methods suffer from the additive bias of the ambient light which has a flattening effect on the reconstruction. There is furthermore a slightly smaller mean normal error than [205] (8.5^o vs 11^o) and this is probably because of the use of the robust variational solver.

The next experiment involves increasing the amount of ambient light (by scaling the image in Figure 4.6), as well as introducing specular reflection, by reducing the specularity parameter c (see Equation 4.3), while keeping the same setup for lighting and orthographic viewing. This is a stress test for the binomial series approximation which displays an increasing inaccuracy with the amount of ambient light(see Section 4.2.3). The comparison is shown in Figure 4.9. As expected, [72, 118] which do not explicitly handle the ambient light, have a steeply increasing MAE as a function of the average level of ambient light. [205] handles the ambient light explicitly and exactly and thus the error barely increases as ambient increases. In contrast, the proposed method has indeed an increase of error as a function of ambient light, although this is insignificant except for the last data point where the ambient light is around 70% of the total image signal. As expected, in the case of specular reflections, the proposed method clearly outperforms all other competing approaches.

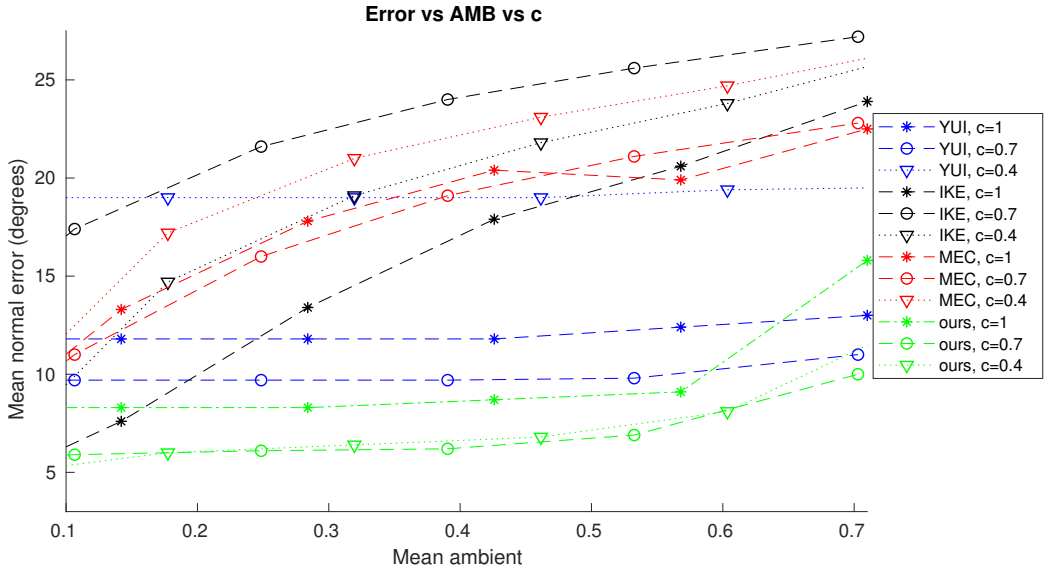


Figure 4.9 Comparison of several PS methods under different ambient light and shininess parameter. The baseline dataset is that of Figure 4.6 but with increasing ambient and specular reflection ($c = 1$ is Lambertian). At high ambient levels, a considerable portion of the pixels get saturated, and this effect dominates the error.

Finally the proposed method is evaluated in the near field scenario (nearby lighting and perspective geometry) with a moderate amount of specular reflection ($c = 0.5$) and with the addition of salt and pepper noise[2]. A few sample images are shown in Figure 4.10 for illustration purposes. Four different levels of noise are considered as well as two different levels of ambient light and the complete results are presented in Table 4.1. Note that evaluation of the proposed method using the Order 2 series approximation is also included. For this experiment it was decided to use a dataset of 9 images only, which allows for consideration of all quadruples of images for the Order 1 method ($\binom{9}{4} = 126$)) and all 6-tuples of images for the Order 2 method ($\binom{9}{6} = \binom{9}{3} = 84$)). The 9 images included 2 circles with 3 sources in the inner and 6 in the outer and the total memory consumption was around 5GB.

As expected, the proposed method (both Order 1 & 2) clearly outperforms all other competing approaches on all experiments. Of greater interest, is a comparison of the Order 1 versus Order 2 versions of the proposed method. As expected the Order 2 approximation obtains a smaller error than Order 1 does in the higher ambient case, but at the same time is more sensitive to noise. This is because of more complicated computation which is more prone to the accumulation of numerical approximations with the difference of images operation amplifying the effect of noise. For the exact same reason both methods proposed

[2]This sets a percentage of pixels to completely black or white and thus models complete outliers to the model.

(a) A=15%,N=0% (b) A=15%,N=3%

(c) A=20%,N=0% (d) A=20%,N=3%

Figure 4.10 Sample images for the synthetic datasets which demonstrate nearby lighting, specular reflection ($c = 0.5$) and salt and pepper noise. The ambient light percentage is with respect to the total image intensity and the noise percentage signifies the expected number of affected pixels.

(Order 1&2) have a higher sensitivity to noise than the other three competitors ([205, 72, 118]). Finally, the difference between Order 1 and 2 is better understood by examining the 3D reconstruction obtained at the top row of Figure 4.11. Order 2 is better at preserving the depth discontinuity at the bottom of the mouth achieving better quality for low frequencies (overall bending) but at the same suffers from a lot of high frequency artefacts that appear like scratches.

Table 4.1 Quantitative evaluation based on the initial estimate quality. Errors are in mm. Noise added to vertex positions and the magnitude is relative to the average triangle size.

		Method-MAE in degrees				
Ambient	Noise	[205]	[72]	[118]	Proposed O1	Proposed O2
15%	0%	19.9	23.8	24.8	9.1	**6.0**
	1%	20.1	23.7	24.9	**11.2**	11.4
	2%	20.3	23.6	24.9	**13.4**	14.7
	3%	20.5	23.5	24.9	**15.7**	16.6
20%	0%	19.9	24.9	28.0	11.1	**7.1**
	1%	20.1	24.9	28.0	13.3	**11.1**
	2%	20.4	24.9	28.0	15.7	**12.5**
	3%	20.6	24.8	28.0	18.1	**14.6**

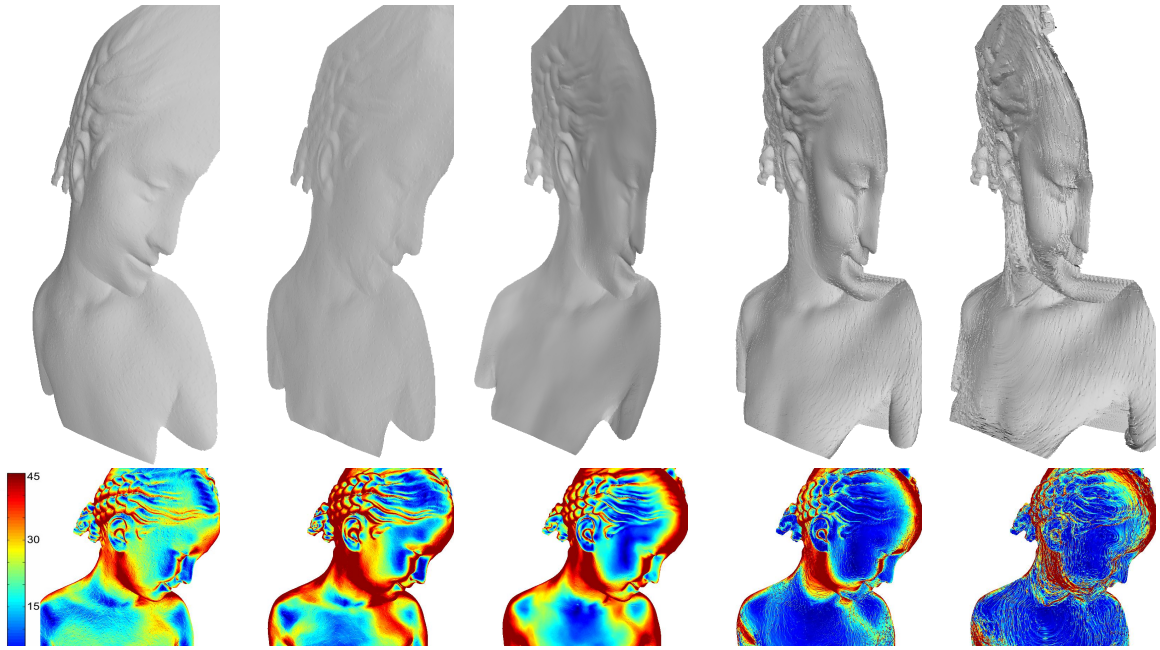
4.4 Discussion

This chapter tackled the problem of PS under ambient light as well as the extensive set of additional realistic assumptions (perspective view geometry, non-linear light propagation, specular reflection) that are common ground for the rest of the thesis. A new approach based on ratios of image differences was presented. This approach is able to remove any additive bias on images. The problem is then expressed as a quasi-linear PDE and is solved through the robust variational optimizer performing ℓ_1 minimization that was introduced in Section 3.7. Experiments on synthetic and real data verify that the proposed approach provides good reconstructions under significant ambient light, specular highlights and perspective deformation.

As a challenge for future work, what can be foreseen is the applicability of this technique involving binomial expansion, from the simple BRDF presented in Section 3.4 to other more physical and complicated BRDFs (see Section 2.1.2). The value of this method is mostly theoretical, showing the ratio of differences method, as well as demonstrating linearising BRDFs through the binomial series expansion method. In practice, it is usually acceptable to capture one image of the ambient light at the beginning of the acquisition and then to simply remove it from the rest of the images.

4.4.1 Limitations

A practical limitation of the work presented so far is that it requires a fully calibrated PS setup, meaning that the light source positions, as well as illumination properties, need to be



(a) [205] MAE= 20.1^o (b) [72] MAE= 23.7^o (c) [118] MAE= 24.9^o (d) Proposed O1 MAE= 11.2^o (e) Proposed O2 MAE= 11.4^o

Figure 4.11 In depth illustration of the 15% ambient light, 1% noise case (row 2 of Table 4.1). The top row shows the reconstructions obtained from all the competing methods and the bottom row shows normal error maps (in degrees). It is noted that the proposed formulation is more sensitive to noise than the other competing approaches, as it is dependent on image differences (which are known to amplify the high frequency components of noise). This is especially true for the Order 2 (O2) formulation which is based on differences of differences, and thus suffers from noise sensitivity the most. Nonetheless, despite the high frequency artefacts, both proposed methods outperform all competitors and generate an overall lower MAE.

known a priori. This assumption will be relaxed in the next chapter where the semi-calibrated PS problem will be introduced and a solution will be proposed.

In addition, the more complicated ambient light PDE parameterisation (Equation 4.10) in contrast to Equation 3.18 means that the problem is more sensitive to numerical inaccuracies. Moreover, as seen in Section 4.2.1 this formulation degenerates under a number of different configurations, in addition to the standard degenerate configurations for PS problems (collinear light sources etc.) Thus, the use of the robust variational solver is essential, as is the preprocessing of data to remove unreliable data points (see Section 3.5.1). An additional remark about unreliable data detection is noted: the standard heuristic of removing dark pixels as an attempt to avoid shadows does not work in the presence of ambient light. Indeed, it is possible to have parts of the surface in the cast shadow region of a light source still

reflecting ambient light. This final issue can be resolved by explicitly detecting surface occlusion by ray-tracing using a previous estimate of the geometry. Indeed, the strategy will be followed in the following chapters in order to minimise the effect of shadows.

Finally, a computational limitation of the proposed method is the fact that the number of n-tuples that need to be considered grows very quickly as a function of the number of images n. This limitation is further discussed in Section 7.2.1 and overcome with a GPU implementation that can very quickly iterate through the required set of image n-tuples.

Chapter 5

Semi-calibrated Photometric Stereo

Up until this point, it has been assumed that the PS setup has been completely calibrated. This chapter extends the PS framework presented in Chapter 3 by relaxing this calibration assumption.

5.1 Motivation

Figure 5.1 The semi-calibrated Photometric Stereo problem: it is very common that different light sources have noticeably different brightness which can be unpredictable, especially for cheap LEDs that tend to flicker. This is demonstrated on these 3 (out of 9) images of a scene consisting of heterogeneous materials.

Photometric Stereo approaches are usually divided into two main categories. The first category solves calibrated PS by requiring an a priori knowledge of the light source's geometry, position and brightness. The second category solves the uncalibrated PS problem where the information about the illumination of the scene is unknown. Recently, Cho et al. [28] introduced the idea of semi-calibrated PS, where the brightness but not the positions of the light sources, was considered unknown. Indeed, accurate knowledge of the brightness, requires careful measurements with either specialized equipment (LUX meters) or reference

objects with known geometry and reflectance. Moreover, the brightness commonly tends to vary from acquisition to acquisition, and this variation is mainly due to three important causes. First of all, the lighting status of a light source: most LEDs used as light sources have a non-negligible ramp-up time when turned on, corresponding to constantly increasing brightness; this is followed by a gradual decrease of brightness because of the heating-up that affects their electrical properties. This is especially relevant for high speed acquisitions performed with high power emitters. Under these circumstances, electrical instabilities causing LED flickering are not uncommon (OSRAM [141]). Secondly, the light sources inevitably deteriorate with time. This imposes the need to perform a calibration of the light sources every time images have to be acquired over a relatively long lapse of time. Finally, it may be desirable to shoot images with auto-exposure settings in order to minimize the number of underexposed and saturated pixels. If this is the case, the effective brightness (i.e. brightness time exposure) of the respective light source cannot be known a priori.

In this chapter, the problem of semi-calibrated near field PS is tackled; this assumes that the geometry of the light sources is reasonably point-wise and their positions are assumed as known. All the realistic modelling assumptions of Chapter 3 are retained including perspective viewing geometry and both diffuse and specular reflectance.

Contribution The robust step-wise variational numerical solver presented in Section 3.7.4 is augmented to alternate between estimating lighting factors, shape and the specularity coefficients. This also allows for the relaxation of the assumption that point light sources spread light uniformly in all directions by computing the spatially varying maps of the light intensities.

The reliability of the proposed method for providing highly accurate 3D reconstructions and approximating lighting parameters is demonstrated quantitatively over a large set of synthetic and real data. Regarding real world experiments, the proposed method is compared against other standard methods for estimating the brightness of light sources (i.e. using a light-meter or reference objects).

5.2 Background

Light calibration, intended as an estimation of light positions/directions and intensities, often requires specific equipment [37, 38] or a dedicated procedure [196] that may differ depending on the scenario. For example, PS for outdoor scenes generally uses intrinsic light calibration according to geo-position information [171, 2, 83, 1] or triangulating with reflective spheres [127]. Alternatively, in a dark environment, light calibration can be performed considering

reference objects where the shape and reflectance are known. Goldman et al. [53] used spheres of the same materials of the object under observation, dealing with spatially varying BRDFs.

5.2.1 Uncalibrated Photometric Stereo

Before dealing with the semi-calibrated problem, the classical uncalibrated problem is worth a mention. As explained in Section 2.2, the classical PS $\boldsymbol{I}_s(u,v) = L_s\boldsymbol{M}(u,v)$ with ($n \geq 3$) calibrated light sources has a unique linear least squares solution $\boldsymbol{M}(u,v) = (L_s^\intercal L_s)^{-1} L_s^\intercal \boldsymbol{I}_s(u,v)$, where $\boldsymbol{M}(u,v) = [\boldsymbol{M}^1, \boldsymbol{M}^2, \boldsymbol{M}^3]^\intercal(u,v)$ is the standard auxiliary variable indicating the albedo-scaled normals. However, if the light source directions are unknown, and the objective is to jointly estimate them along with the surface geometry, an ambiguity arises as

$$\boldsymbol{I}_s(u,v) = L_s\boldsymbol{M}(u,v) \Longleftrightarrow \boldsymbol{I}_s(u,v) = (L_sT)(T^{-1}\boldsymbol{M}(u,v)) \tag{5.1}$$

for any invertible 3x3 matrix T. As shown by Belhumeur et al. [14], this 9 parameter ambiguity can be reduced by employing the integrability constraint (see Section 2.2.2). Indeed, starting from

$$\frac{\partial(\frac{\boldsymbol{M}^2}{\boldsymbol{M}^3})}{\partial u} = \frac{\partial(\frac{\boldsymbol{M}^1}{\boldsymbol{M}^3})}{\partial v} \tag{5.2}$$

and after some algebra, it was proved that the only compatible transformation is the *Generalized Bas-Relief* (GBR) which has the form

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix} \tag{5.3}$$

where μ, ν and λ are the three parameters of the ambiguity. Finally, it was shown that if the albedo is known or constant (or equivalently, the light sources have the same magnitude) then the ambiguity reduces to binary convex/concave: $\mu = 0$, $\nu = 0$ and $\lambda \pm 1$

The inherent ambiguity in uncalibrated PS disappears when assuming perspective instead of orthographic viewing as shown by Papadhimitri and Favaro [142]. That is because under perspective projection (see parameterisation in Section 3.3.1) the normal depends on the depth non-linearly and this non-linearity combined with the integrability constraint breaks the linear GBR ambiguity. In addition, their work is extended to uncalibrated PS with point light sources in Papadhimitri and Favaro [143].

Other uncalibrated approaches include Quéau et al. [151] which requires Dichlet boundary condition for the depth, Shi et al. [167] who disambiguate the (GBR) ambiguity through

analyzing color/intensity profiles in the RGB and irradiance-time domains, Favaro and Papadhimitri [46] who exploit points of maximum diffuse reflection and finally Georghiades [51] who overcomes the ambiguity up to binary convex/concave through the use of the non-linear Torrance and Sparrow reflectance model [176].

5.3 Semi-calibrated classical Photometric Stereo

First of all, a quick review of the method of Cho et al. [28] is presented as this is the foundation for this chapter. This involves extending the Classic PS method of Woodham [192] (see Section 2.2) by including an additional unknown per image, ϕ_j that denotes the light source brightness. Therefore, the irradiance Equation 2.11 is extended to:

$$i_j(u,v) = \phi_j \rho(u,v)(\overline{\boldsymbol{N}}(u,v) \cdot \overline{\boldsymbol{L}}_j) = \phi_j \overline{\boldsymbol{L}}_j^{\intercal} \boldsymbol{M}(u,v) \tag{5.4}$$

The number of different light sources is n and the number of pixels is p. Stacking all equations from all pixels and all light sources into a combined linear system gives:

$$I = \begin{bmatrix} \phi_1 & 0 & \dots \\ 0 & \phi_2 & \dots \\ \vdots & & \\ \dots & 0 & \phi_n \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{L}}_1^{\intercal} \\ \overline{\boldsymbol{L}}_2^{\intercal} \\ \vdots \\ \overline{\boldsymbol{L}}_n^{\intercal} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_1^1 & \boldsymbol{M}_2^1 \dots \boldsymbol{M}_p^1 \\ \boldsymbol{M}_1^2 & \boldsymbol{M}_2^2 \dots \boldsymbol{M}_p^2 \\ \boldsymbol{M}_1^3 & \boldsymbol{M}_2^3 \dots \boldsymbol{M}_p^3 \end{bmatrix} = \Phi L M \Leftrightarrow \Phi^{-1} I - LM = \mathbf{0} \tag{5.5}$$

where in Equation 5.5 I is a $n \times P$ matrix stacking all data from all images (noting that the matrix Φ is diagonal and hence invertible). Then, Equation 5.5 can be solved using the following vectorisations:

$$LM = L \begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \dots \boldsymbol{M}_p \end{bmatrix} \rightarrow \begin{bmatrix} L & 0 \dots 0 \\ 0 & L \dots 0 \\ & \vdots & \\ 0 & 0 \dots L \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{n \times p} \end{bmatrix} = (J_p \otimes L)\text{vec}(M) \tag{5.6}$$

and

$$\Phi^{-1} I = \begin{bmatrix} 1/\phi_1 & 0 & \dots \\ 0 & 1/\phi_2 \dots 0 & \\ & \vdots & \\ 0 & 0 \dots 1/\phi_k & \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_1 & \boldsymbol{I}_2 \dots \boldsymbol{I}_P \end{bmatrix} \rightarrow \begin{bmatrix} \text{diag}(\boldsymbol{I}_1) \\ \text{diag}(\boldsymbol{I}_2) \\ \vdots \\ \text{diag}(\boldsymbol{I}_P) \end{bmatrix} \begin{bmatrix} 1/\phi_1 \\ 1/\phi_2 \\ \vdots \\ 1/\phi_k \end{bmatrix} = I_d \boldsymbol{\psi} \tag{5.7}$$

where in Equation 5.6 J_p is an identity matrix of size p, and $\otimes$ and vec are Kronecker product and vectorisation operators respectively. In Equation 5.7, diag is the diagonalisation operator. Combining Equations 5.6 and 5.7 gives:

$$I_d\boldsymbol{\psi} - (J_P \otimes L)\text{vec}(M) = \mathbf{0} \rightarrow \underbrace{\begin{bmatrix} I_d\boldsymbol{\psi} & -(J_P \otimes L) \end{bmatrix}}_{E} \begin{bmatrix} \boldsymbol{\psi} \\ \text{vec}(M) \end{bmatrix} = \mathbf{0} \tag{5.8}$$

with E being a $np \times (3p + n)$ sparse matrix. The homogeneous Equation 5.8 has a non-trivial, up to scale solution, if the rank of the matrix E is $3p + n - 1$ and thus E has a one dimensional null space. In general, under noise and other imperfections, the least squares solution is found by computing the SVD of E and selecting the basis vector corresponding to the smallest singular value (which will not be exactly 0).

This is a remarkable result showing that the semi-calibrated PS problem is well-posed and does not suffer from ambiguities like that of the Generalised Bass Relief (see Section 5.2.1), even when keeping the classic PS setup. However, directly solving Equation 5.8 is completely impractical, as E is too big for one to be able to compute its SVD. To overcome this limitation, Cho et al. [28] proposed an alternating optimisation scheme, where M and $\boldsymbol{\Phi}$ are linearly estimated using each others previous estimates.

5.4 Semi-calibrated near-field Photometric Stereo

After examining the baseline method of Cho et al. [28], this section shows how it can be extended to meet the realistic modelling assumptions required, to work in the near field. The key idea is to extend their alternating optimisation scheme in order to calculate all the required quantities including albedo, light source brightness and depth. Thus, the overall irradiance equation for the j^{th} light source is:

$$i_j(u, v) = \phi_j \rho(u, v) a_j(u, v, z(u, v)) \left(\overline{\boldsymbol{N}}(u, v, z(u, v)) \cdot \overline{\boldsymbol{W}}_j(u, v, z(u, v), c) \right)^{\frac{1}{c(u,v)}}. \tag{5.9}$$

Equation 5.9 is almost identical to Equation 3.16. For now, it is assumed that the light attenuation follows the point-wise anisotropic model of $a(u, v)_j = (\overline{\boldsymbol{P}_j} \cdot \overline{\boldsymbol{L}_j(u, v)})^{\mu}$ with $\boldsymbol{P}_j$ the principal direction of the light source and μ the attenuation coefficient. It is noted that assuming ϕ_j is known, then the problem reduces to the standard near field photometric stereo which can be solved using the variational optimisation scheme described in Algorithm 1 in Section 3.7.6.

The rest of the chapter explains how to calculate the rest of the relevant quantities as well as how to perform an alternating optimisation scheme that combines everything together.

5.4.1 Albedo and light source brightness

The problems of computing the albedo, the light source brightness and the light attenuation maps are formulated such that they can be directly computed depending on the geometry of the scene z and the reflectance coefficient $c(u, v)$ of the irradiance Equation 5.9.

First of all, normals can be computed by numerically differentiating the depth. The weighted lighting vector $\boldsymbol{W}_j(u, v, z(u, v))$ is also trivially computed using the current estimate of the depth and the light source positions.

Rearranging the irradiance Equation 5.9 gives the following relation for the albedo:

$$\rho(u, v) = \frac{i_j(u, v)}{\phi_j a_j(u, v, z) d_j(u, v, z)} \tag{5.10}$$

where $d_j(u, v, z) = (\overline{\boldsymbol{N}}(u, v) \cdot \overline{\boldsymbol{W}}_j(u, v, z))^{\frac{1}{c(u,v)}}$.

We note that for each pixel (u, v), there are n equations (5.10). All of them can be stacked into a matrix giving the following (over-constrained) linear system:

$$\begin{bmatrix} \phi_1 a_1(u, v, z) d_1(u, v, z) \\ \vdots \\ \phi_n a_n(u, v, z) d_n(u, v, z) \end{bmatrix} \rho(u, v) = \begin{bmatrix} i_1(u, v) \\ \vdots \\ i_n(u, v) \end{bmatrix}. \tag{5.11}$$

For the light source brightness, the equation (5.10) can be equivalently re-arranged as:

$$\phi_j = \rho(u, v) \frac{i_j(u, v)}{a_j(u, v, z) d_j(x, y, z)}. \tag{5.12}$$

It is noted that for each light source n there are p equations (5.12), one per each pixel ordered as $(u_1, u_1), \ldots, (u_p, v_p)$. Stacking these into a linear system, gives:

$$\begin{bmatrix} \rho(u_1, v_1) a_j(u_1, v_1, z_1) d_j(u_1, v_1, z_1) \\ \vdots \\ \rho(u_p, v_p) a_j(u_p, v_p, z_p) d_j(u_p, v_p, z_p) \end{bmatrix} \phi_j = \begin{bmatrix} i_j(u_1, v_1) \\ \vdots \\ i_j(u_p, v_p) \end{bmatrix}. \tag{5.13}$$

(a) Sample Image (b) Shadow map (c) Point a (d) Calculated a

Figure 5.2 (d) shows a sample calculated attenuation map $a(u, v)$ of one of the images (shown in (a)) from the sequence of Figure 5.1. (b) shows the calculated shadow map (see Section 5.5.1) which in turn makes the calculation of attenuations for the shaded pixel irrelevant, and thus these are left white in (d). (c) shows the attenuation map obtained using the standard point light source assumption, and is included here as a comparison to (d).

5.4.2 Attenuation map

Most near field PS methods [74, 30, 201] assume point light sources with uniform radial spread (see Section 3.3.2) which leads to an inverse square attenuation term. It is also common to assume an angular anisotropy factor which has to be calculated from the LED data-sheet. However, in practice this assumption might not be very realistic, for a number of reasons, for example, because the light source has a finite size and is not a perfect point. In addition, it may be hard to get an estimate of the anisotropy parameter. Finally, there is some interest in the literature to perform PS in attenuating media which are even harder to model analytically (e.g. in murky water Murez et al. [126]) . Ideally, the light attenuation at each pixel for each image should be calculated by considering $a_j(u, v)$ as a pixel-wise unknown. However, this will render the PS problem "hopelessly under-constrained" (see Section 2.2.1) with many more unknowns than equations. To overcome this limitation, the assumption that the light attenuation $a_j(u, v)$ is locally constant is employed. Note that this is a much weaker assumption than assuming that the lighting is locally directional. Indeed, $\boldsymbol{L}_k(u, v)$ is expected to have a non-negligible variation even between nearby pixels.

Thus, what is sought here, is to solve for $a_j(u, v)$ by assuming that it has a constant value over a small patch of k pixels surrounding each pixel (u, v). For every patch, each pixel in each image provides one equations, thus the total number of equations is $n \times k$. The unknowns to be calculated are: k albedos $\rho(u, v)$, k depths $z(u, v)$, k shininess coefficients $c(u, v)$ and n attenuation $a_j(u, v)$.

Thus, in order for the number of equations to exceed the number of unknowns, it is required that:

$$n \times k > 3k + n \iff k > \frac{n}{n-3}. \tag{5.14}$$

The inequality (5.14) is easily satisfied for $k = 9$ (i.e. a 3x3 pixel patch) and $5 \lesssim n \lesssim 50$ (typical number of images used in PS problems). Under this assumption establishing that the search for $a_k(u, v)$ is not under-constrained, Equation 5.9 is re-arranged as:

$$a_j(u, v) = \frac{i_j(u, v)}{\rho(u, v)\phi_j d_j(u, v)}. \tag{5.15}$$

Reusing the strategy followed in Section 5.4.1, it is noted that for each light source j and each patch, there are k equations (5.15), one per each pixel ordered as $\Big((u_1, v_1), \ldots, (u_k, v_k)\Big)$. Stacking these into a linear system gives:

$$\begin{bmatrix} \rho(u_1, v_1)\phi_j d_j(u_1, v_1) \\ \vdots \\ \rho(u_k, v_k)\phi_k d_j(u_k, v_k) \end{bmatrix} a_j(u, v) = \begin{bmatrix} i_j(u_1, v_1) \\ \vdots \\ i_(u_k, v_k) \end{bmatrix}. \tag{5.16}$$

5.4.3 Scale ambiguity

There is one issue that arises from the above-mentioned discussion, and that is, a global two parameter ambiguity between the scale of ϕ_j, the scale of $a_j(u, v)$ and the scale of $\rho(u, v)$. Indeed, one can replace ϕ_j by $\mu\,\phi_j$, $a_j(u, v)$ by $\nu\,a_j(u, v)$ and $\rho(u, v)$ by $\frac{\rho(u,v)}{\mu\nu}$, for any scalars μ, ν, and the irradiance Equation 5.9 remains unchanged.

To overcome these ambiguities, the following conventions are used: $\|\mathbf{\Phi}\| = 1$, with $\mathbf{\Phi} = [\phi_1, \ldots, \phi_n]^\intercal$ and also mean$[\rho(u, v)] = 1$. Although this may seem counter-intuitive with respect to the usual definition of $\rho(u, v)$ to always be between 0 and 1, the irradiance equation is consistent with $\rho(u, v)$ of any scale.

5.5 Computational approach

In Section 5.4, it was shown how the calculation of the various unknowns for the semi-calibrated near field PS problem can be cast as a series of linear (over-constrained) problems expressed in Equations 5.11, 5.13 and 5.16. Subsequently, inspired by Cho et al. [28] (see Section 5.3), an alternating optimisation strategy is employed to solve the overall problem. First of all, it is noted that assuming light source brightness $\mathbf{\Phi}$ and attenuation $a(u, v)$ as known, the depth $z(u, v)$ and specularity parameter $cz(u, v)$ calculation steps reduce into the standard calibrated near field PS and can be solved using the variational optimisation strategy of Section 3.7.

It is noted that the $\mathbf{\Phi}$, $\rho(u, v)$ and $a(u, v)$ update steps come down to solving a series of linear systems of the form $A_s\boldsymbol{X} = \boldsymbol{B}_s$. As these are over-constrained, there is not only

one solution, and to maximise robustness, the next step is to find the ℓ_1 solution, namely $\min_{\boldsymbol{X}} \|A_s\boldsymbol{X} - \boldsymbol{B}_s\|_{\ell_1}$. This could be achieved using the ADMM scheme of Section 3.7.4 (which is used for the computation of the depth). However this could be computationally expensive since the albedo calculation, needs to run once per pixel. A simpler solution (from Candès et al. [23]) based on re-weighted least squares (wLSQ) is shown in Algorithm 2:

Algorithm 2: Iteratively re-weighted least squares ℓ_1 solver

Input: Matrix A_s, vector $\boldsymbol{B}_s$
Output: vector $\boldsymbol{X}$ such as $\min_{\boldsymbol{X}} \|A_s\boldsymbol{X} - \boldsymbol{B}_s\|_{\ell_1}$
while $\|A_s\boldsymbol{x}^t - \boldsymbol{B}_s\|_{\ell_1} < tol$ **do**
  calculate residual: $\boldsymbol{R} = A_s\boldsymbol{X}^t - \boldsymbol{B}_s$;
  set $w_i = \frac{1}{\epsilon + \|\boldsymbol{R}_i\|_{\ell_1}}$;
  set $W = \text{diag}(\boldsymbol{W})$;
  solve wLSQ: $\boldsymbol{X}^{t+1} = (A_s^\intercal W A_s)^{-1} A_s^\intercal W \boldsymbol{B}_s$;
end

In Algorithm 2 the constant ϵ ensures numerical stability, and [23] recommend setting it at a magnitude slightly smaller than the expected magnitude of the vector components, so it is set to $1/\text{size}(\boldsymbol{X})$.

5.5.1 Pixel based selection strategy

As the aim is to solve for a large number of unknowns, it is imperative to minimise the effect of outliers. This is somewhat solved with the use of the ℓ_1 error above. In addition, a dynamic selection of reliable pixels is employed so as to eliminate numerically unstable data. The following set of heuristics are used (it is assumed that image values are normalised with 0 corresponding to black, 1 corresponding to white):

- Very low image values ($i_j < 0.05$) as these break the ratio method.
- Saturated image values ($i_j > 99$).
- Attached shadows or very little illumination ($\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{L}}_j < 0.1$).
- Cast shadows using ray-tracing (see below).

Shadow map estimation For the near field scenario, when light sources are close to the object, shadows appear frequently. There are several works dealing with shadows, aiming to extract geometrical information [172] or to avoid biased shading information [60] especially

when only a few light sources are considered. For example, Chandraker et al [26] proposed a graph-cut method to estimate the visible pixels. Barsky and Petrou [10] took into account both highlights and shadows using only four light sources. However, different approaches for 3D shape recovery such as multi-view stereo techniques, including [184, 206], have employed direct visibility computation steps. This has the potential of higher reliability than the heuristics presented above.

Since an iterative, global geometry refinement algorithm is followed, it is reasonable to calculate cast shadows directly, through ray-tracing using the previous estimate of the geometry. Although this procedure is very expensive in terms of computational time, it ensures that the surface is globally consistent with the light sources making the computation of the lighting factors more reliable.

A more computationally efficient way of performing this ray-tracing operation, through the use of an octree data structure, will be discussed in Section 6.5.2.

5.5.2 Alternating optimisation

Up to this point, the discussion has centred on how different sets of variables can be used to calculate the rest. Of course, when first approaching the problem, very little information about the geometry and the photometric properties of the scene is known and all the unknowns have to be jointly estimated. In addition, as each step depends on the results of the previous steps, the order of these steps is crucial so as the overall procedure to converge to an acceptable solution at a reasonable rate.

First of all, the $\mathbf{\Phi}$ update step is performed according to (5.13). A single scalar per image is calculated using the whole image data, thus this step is very robust to outliers and errors. The dependence of the system (5.13) on z is overcome by initialising $z^0 = z_{\text{mean}}$, the mean distance between the camera and the object. The albedo is initialized to $\rho^0(u, v) \equiv 1$, the attenuation using the point source model $a_j^0(u, v) = \frac{1}{\|\boldsymbol{L}_j(u,v,z^0)\|^2}$ and $\phi_j = \frac{1}{\sqrt{n}}, \forall j$. Finally, c is initialised as $c(u, v) = 1$, which corresponds to diffuse reflectance.

The next step updates the z values using the variational solver (Section 3.7) which is very robust to noise and other forms of sparse corruptions but cannot deal with the systematic error coming from using inaccurate $\mathbf{\Phi}$. Indeed, this is exactly the core purpose of the whole chapter; inaccurate estimates for $\mathbf{\Phi}$ can lead to substantial deformation of the recovered shape. This is demonstrated experimentally in Section 5.6 in the comparison with the baseline calibrated near-field PS (Chapter 3). Furthermore, the main equation for this step Equation 3.18, is albedo independent hence it is reasonable to precede this computation to the computation of the albedo.

Algorithm 3: Semi-calibrated PS

Input: Images, light source positions, rough mean distance z_0

Output: Depth map $z(u,v)$, light source brightnesses $\mathbf{\Phi}$, aldebo $\rho(u,v)$, attenuation maps $a_j(u,v)$ and shininess parameter $c(u,v)$

Initialisation:

$z^0(u,v) = z_0$, $c^0(u,v) = 1$, $\rho^0(u,v) = 1$, $\phi_j = \frac{1}{\sqrt{n}} \forall j$;

calculate fields $\boldsymbol{L}_j^0(u,v,z^0) \quad \forall j$;

set $a_j^0(u,v,z) = \frac{1}{\|\boldsymbol{L}_j^0(u,v,z^0)\|^2} \quad \forall j$;

while $\mid z^{t+1} - z^t \mid > 10^{-4} \times \mid z^t \mid$ **do**

  if $t > 1$ compute shadow maps. **endif** ;

  calculate fields $\boldsymbol{W}_j^t(u,v,z^t) \quad \forall j$;

  calculate $\boldsymbol{N}^t(z^t)$ (finite differences);

  compute $\mathbf{\Phi}^{t+1}$ solving (5.13);

  for every pair j and k, calculate fields $\mathbf{b}_{jk}(z^t, a_j^t, a_k^t, \mathbf{\Phi}^{t+1}, \mathbf{W}_j^t, \mathbf{W}_k^t)$, $s_{jk}(z^t, a_k^t, a_k^t, \mathbf{\Phi}^{t+1}, \mathbf{W}_j^t, \mathbf{W}_k^t)$;

  calculate z^{t+1} (§ 3.7) ;

  calculate $\boldsymbol{N}^{t+1}(z^{t+1})$ (finite differences);

  calculate fields $\boldsymbol{W}_j^{t+1}(u,v,z^{t+1}) \quad \forall j$;

  $\rho^{t+1}(u,v)$ solving Equation 5.11;

  $a_j^{t+1}(u,v,z^{t+1})$ solving Equation 5.16 $\forall j$;

  calculate c^{t+1} (§ 3.7.5) ;

end

The next step is the attenuation map calculation. This is a local operation and much more sensitive to inaccuracies. For each pixel, $3 \times 3 = 9 = k$ sized patches containing all of their neighbours, are chosen to estimate the attenuation. As the patches are overlapping, continuity is implicitly enforced.

Finally, the $c(u,v)$ update is performed at the end. This is the most numerically unstable step as it involves a ratio of differences of logarithms thus it is essential to have as good an estimate of the rest of the parameters as possible.

The whole alternating optimisation procedure is summarized in Algorithm 3.

5.5.3 A non-ratio alternating optimisation approach

Concurrently with our work in [103], Quéau et al. [153] presented another semi-calibrated, variational approach, which is also included in their extensive study of near-field methods in [149]. They have a similar aim of modelling nearby, point light sources (using the same

parameterisation as the one presented in Section 3.3.4), as well as perspective projection, but differ in that they assume Lambertian reflection. Their main difference with the proposed approach is that they avoid considering image ratios and instead eliminate the non-linearity at the normalisation of the normal by handling it as an auxiliary variable. In addition, the division by z of the perspective projection is removed, using the standard $g = \log(z)$ transformation (see Section 3.3.1). Thus they obtain a quasi-linear PDE which is solved in an alternating optimisation procedure, also incorporating a robust estimator.

Thus, in the next section, comparison with this method is also included.

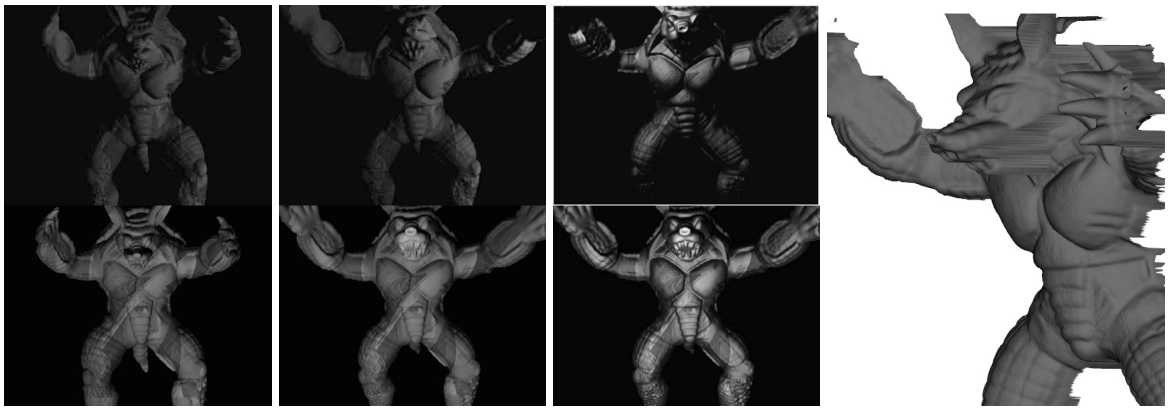
5.6 Experiments

The proposed algorithm is evaluated on synthetic and real data which cover a range of different situations. Firstly, in order to make the proposed approach comparable to the state-of-the-art semi-calibrated PS Cho et al. [28] (by using a simple implementation of their alternating minimisation method), Algorithm 3 was adapted to the easier, fully orthographic and diffuse scenario. This was easily achieved by fixating $c = 1$, $a_j = 1$ and $f = \infty$ (10^6 in practice). Secondly, in order to test the reliability of the proposed approach, synthetic data generated with the Cook and Torrance reflection model (see section 2.1.2) were considered. Furthermore, comparison with the baseline calibrated near-field Photometric Stereo (Chapter 3) is provided. The comparison with this work provides clues regarding shape deformation which occurs when reasonable, but incorrect brightness is used. Finally, some comparison with the later semi-calibrated method Quéau et al. [149] is also provided.

The proposed algorithm was implemented in MATLAB and run on a server machine with an AMD Opteron CPU. The computation time for the 2 MPixel x 9 images datasets, was about one hour, with peak memory usage of around 30GB. Around 90% of the total computation time was spent on the shadow and attenuation maps calculation steps.

5.6.1 Synthetic data

Three synthetic data sets, with eight images each, were generated, using the “Armadillo” model (see Figure 3.6) with the “Lena” image was used for the albedo (see Figure 5.3). The first dataset was made under the classic PS assumptions i.e. directional lighting, orthographic projection and diffuse reflection, in order to provide a fair comparison against Cho et al. [28]. The second dataset was a near-field, diffuse reflection scenario where the object had a size of 2cm and was placed 4cm away from a virtual pinhole which had a focal length of 12mm. The light sources were symmetrically distributed at two circles of radius 3cm and

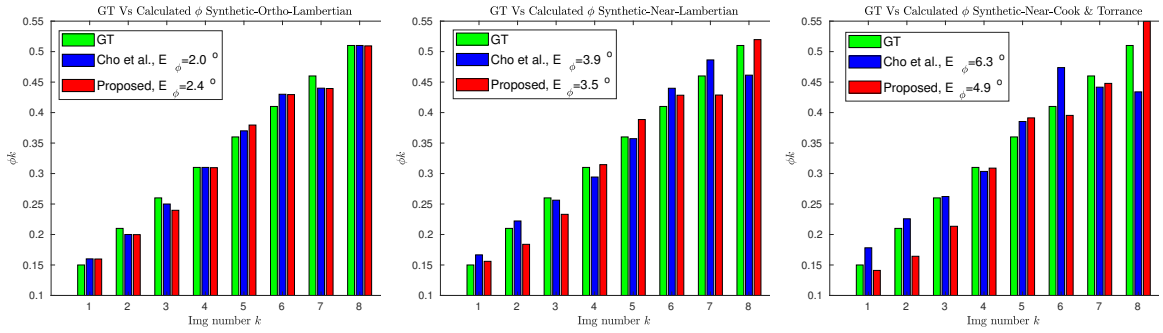


(a) (b) (c) (d) Ground truth

Figure 5.3 Synthetic data samples and ground truth mesh (d). (a), (b) and (c) correspond to the three experiments performed namely classic PS, near-field with Lambertian reflection and near-field with Cook & Torrance reflection respectively.

5cm respectively, around the camera. To apply Cho et al. [28] to the near-field scenario, the $\boldsymbol{L}_j$ at the center of the object was used as the lighting direction. As a result, the errors grew towards the boundaries of the image (Figure 5.5). Finally, the third dataset had the same arrangement as the near field scenario but the object was rendered by using the Cook and Torrance reflection model. This created specular highlights inconsistent with the assumed irradiance Equation 5.9, thus the robustness of the proposed method was tested with more physically derived synthetic data. To stress all compared algorithms to their limits, the brightness of the light sources $\boldsymbol{\Phi}$ was made to vary significantly. In fact, the brightest light source was 5 times brighter than the dimmest (see Figure 5.4).

The quantitative evaluation of the algorithms was achieved through the Mean Angular Error (see Section 4.3.1) between the normals obtained and the ground truth, shown in the bottom line of Figure 5.5. The evaluation of the predicted light source intensities is shown in Figure 5.4. It was achieved by finding the angle between $\boldsymbol{\Phi}$ and the ground truth $\boldsymbol{\Phi}_{gt}$. As $\boldsymbol{\Phi}$ is always a unit vector (section 5.4.3), $E_{\boldsymbol{\Phi}} = \arccos(\boldsymbol{\Phi} \cdot \boldsymbol{\Phi}_{gt})$. This angle is reported in degrees although, as these are 9 dimensional vectors (for datasets with 9 images), the reported values have limited physical meaning. The proposed approach significantly outperforms the other approaches being challenged in both near field cases and gets a slightly higher error than [28] in the classic PS scenario. This final result is expected, as [28] is designed to solve the PS problem when the classic PS assumptions are exactly true.



(a) Classic PS (b) Near/Diffuse. (c) Near/Cook & Torrance.

Figure 5.4 Quantitative evaluation of the light source brightness estimation. The comparison is against [28] which was at the time the only other semi-calibrated approach. The evaluation metric is the angle between estimated $\mathbf{\Phi}$ and the ground truth. The proposed approach is marginally worse at the classic PS scenario (a) but outperforms it in the cases of near field lighting and specular reflectance.

5.6.2 Real data

The proposed algorithm was evaluated in several real datasets shown in Figure 5.6 as well as the dataset shown in Figure 5.1. To make the tests as challenging as possible, multiple objects with various reflections were combined, including a plastic baseball player figurine, a marble statue, a 3D printed plastic version of the Armadillo and a shell. The algorithm was initialised to a flat plane at a distance approximated by a ruler measurement. For scenes which are roughly 10-20cm away from the camera, the 1mm limit of precision of the ruler, translates to around 0.5-1% uncertainty. To simulate significant variation in light source brightness, in order to stretch the algorithm to its full potential, images in Figure 5.1 were shot with varying exposure.

Quantitative Evaluation In order to quantitatively evaluate the semi-calibrated PS method, estimates of the varied brightness of the light sources (contained in the unit vector $\mathbf{\Phi}$) are compared with the ground truth in the same way as for the synthetic experiments presented in Section 5.6.1. The ground truth light source brightness is calculated using two different approaches. The first approach involved using a LUX meter, by placing its sensor in the middle of the field of view at a known distance. Thus, $\mathbf{\Phi}$ can be calculated by multiplying the LUX value with the distance attenuation factor (see Section 3.3.4) which is calculated precisely[1] since the LUX meter sensor is in a know position. In addition, very slow acquisition sequences were performed where each LED was given around 100ms to reach the end of its

[1]assuming the angular anisotropy factor is accurate.

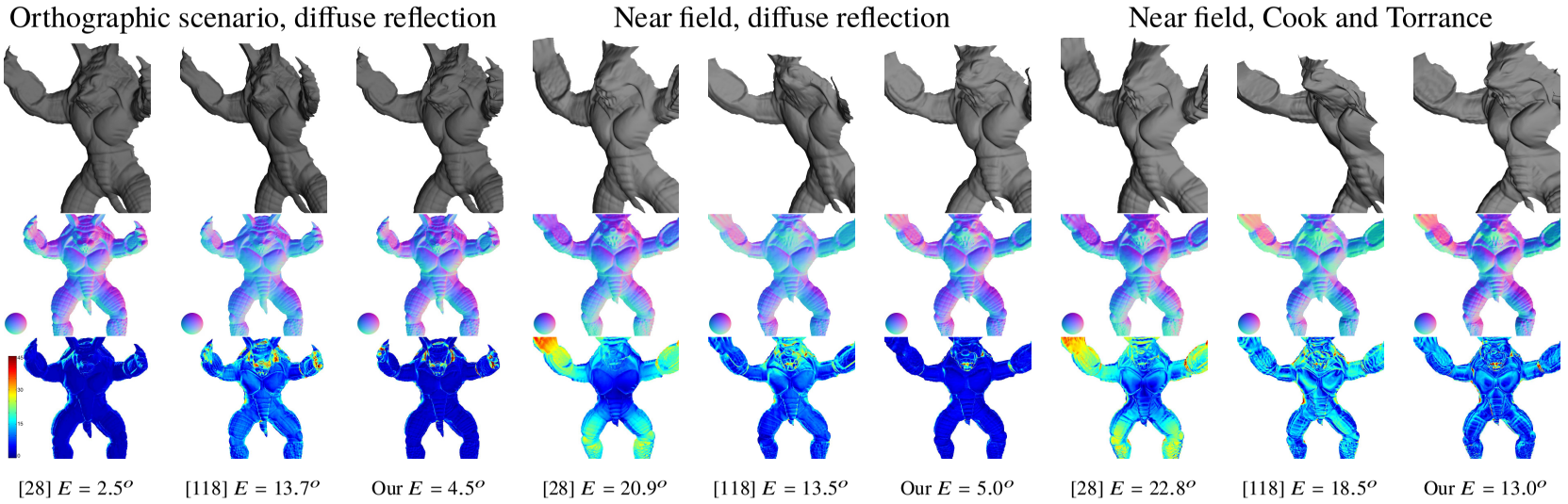


Figure 5.5 Comparison with the semi-calibrated classic PS (Cho et al. [28]) and the baseline calibrated method Mecca et al. [118] (see Chapter 3). The three columns correspond to increasingly challenging experiments: semi-calibrated classic PS (*left*) (see Section 5.3), nearby light sources with perspective projection and diffuse reflection (*center*) and finally, the near-field setup with Cook and Torrance reflectance (*right*). The images presented here include (from top line to bottom): reconstructions, normal maps, and angular error for normals compared to the ground truth. The proposed approach outperforms [28] in the two near-field experiments and it is worse only in the classical PS scenario, as this is accurately modelled by [28] while all of the additional modelling/robustness of the proposed approach is irrelevant. As expected, the baseline calibrated method Mecca et al. [118] is outclassed.

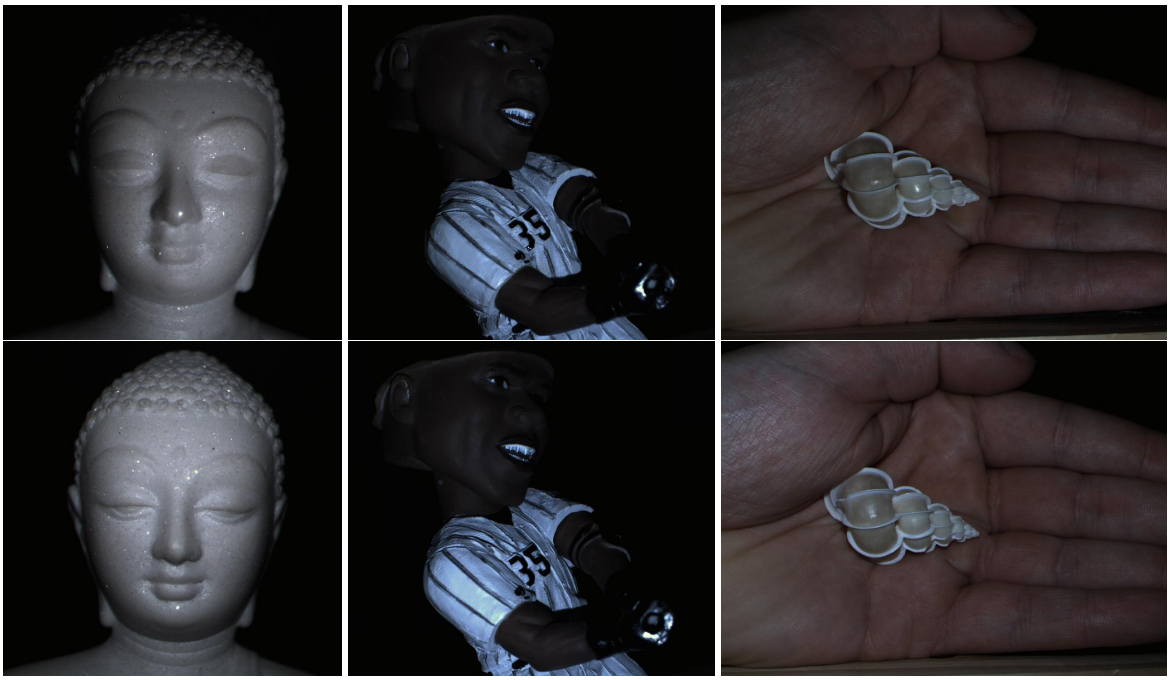


Figure 5.6 Two sample images for each of the three additional real datasets (along with the one presented in Figure 5.1). The top line contains the darkest image of each dataset while the bottom line has the brightest image of each dataset.

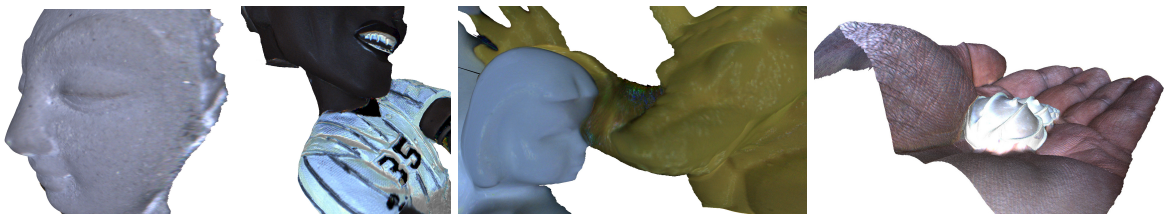
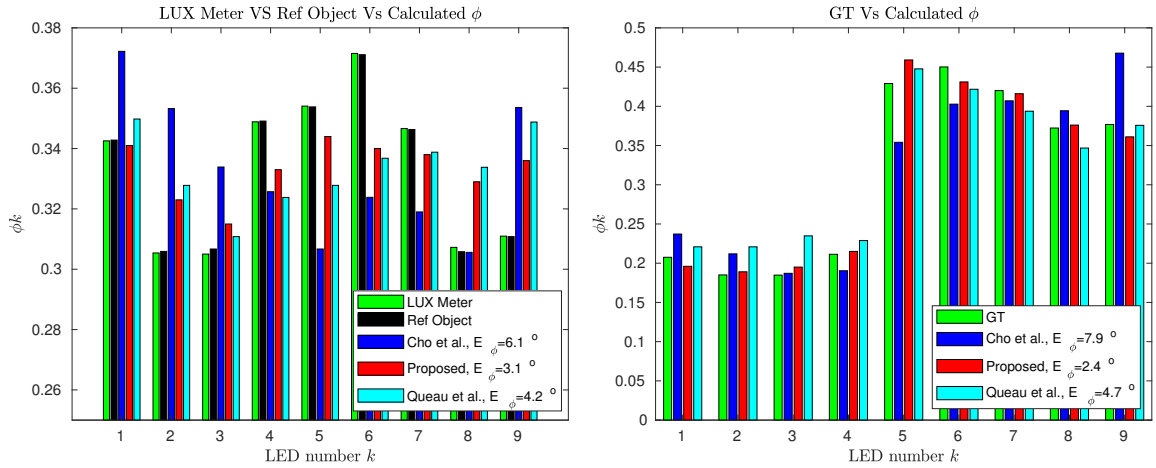


Figure 5.7 Close-up views of the textured reconstructions of the datasets of Figures 5.1 and 5.6, obtained with the use of the proposed method.

rump-up time before acquiring images. A delay of a few seconds was introduced between capturing from different LEDs, to avoid overheating.

The second approach for the calculation of the ground truth Φ was image based: For each LED, a picture of a planar surface, perfectly aligned to the image plane and painted with a water based barium sulfate coating formulated to yield paint with high diffuse reflection, was taken Then, the $\mathbf{\Phi}$ estimation was done by ignoring the z and c update steps of the proposed algorithm (known geometry & reflectance on the reference object). This essentially meant alternating between solving Equations 5.11 and 5.13 which led to a very quick convergence. The second method was intended to be a sanity check, and the fact that it lead to an almost identical estimate for $\mathbf{\Phi}$ (see Figure 5.8a) confirms the reliability of that estimate.

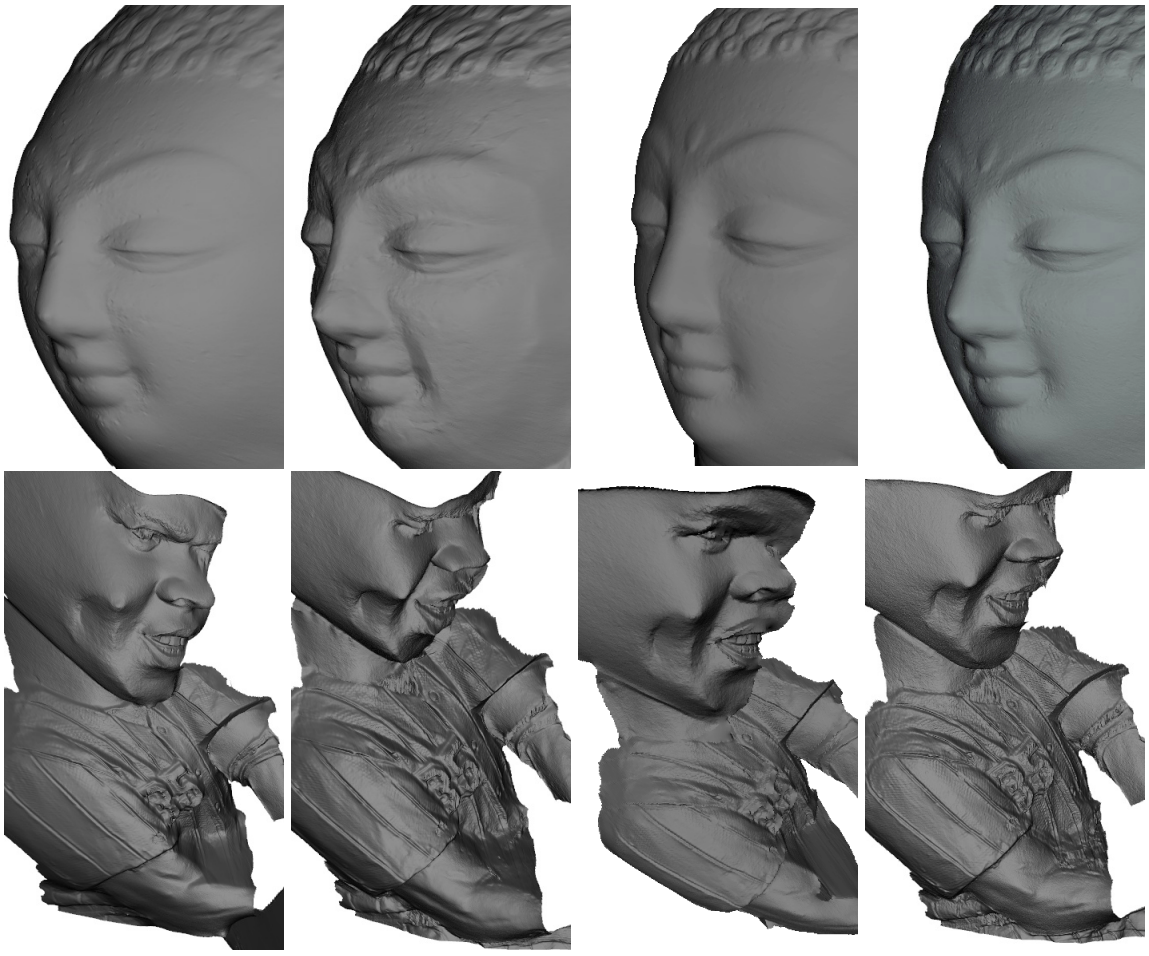


(a) Hand Dataset. (b) Mother-child-Armadillo dataset.

Figure 5.8 Quantitative evaluation on light source brightness estimation. The ground truth $\mathbf{\Phi}$ is provided by multiplying the LUX meter luminance estimate with the known exposure value (as the dataset of (b) is shot with variable exposure in order to increase the challenge of the dataset). In addition, a reference object (white-Lambertian plane) is used as a sanity check and provides a very similar $\mathbf{\Phi}$ estimate as the LUX meter. The comparison is performed with the other two semi-calibrated methods Cho et al. [28], Quéau et al. [149] (for an explanation of [149], see Section 5.5.3) and the proposed approach outperforms both of them.

Regarding the result of $\mathbf{\Phi}$ estimation as shown in Figure 5.8 note that the proposed approach outperforms both competing semi-calibrated methods ([28, 149]), especially on the highly specular "Mother-child-Armadillo" dataset (Figure 5.8b). This is expected as the proposed method specifically handles specular reflections in contrast to [28] which does not handle them at all and [149] which indirectly deals with them with a robust estimator.

Qualitative Evaluation A qualitative comparison between the proposed method and the competing methods is shown in Figures 5.9, 5.10 and 5.11 . As expected, the baseline calibrated near field method ([118]) is entirely outclassed and this demonstrates the importance of the semi-calibrated problem. In addition, the reconstructions of [28] experience some deformations due to near field effects and specular reflections. Finally, in terms of a qualitative assessment, the results of the proposed method and those of [149] are very similar, especially in the almost Lambertian Buddha dataset (5.9), and indeed, this shows the value of incorporating a robust estimator. Nonetheless, the quantitative metric of light source brightness reveals a slight advantage of the proposed method over the other methods, especially as far as the specular datasets are concerned.



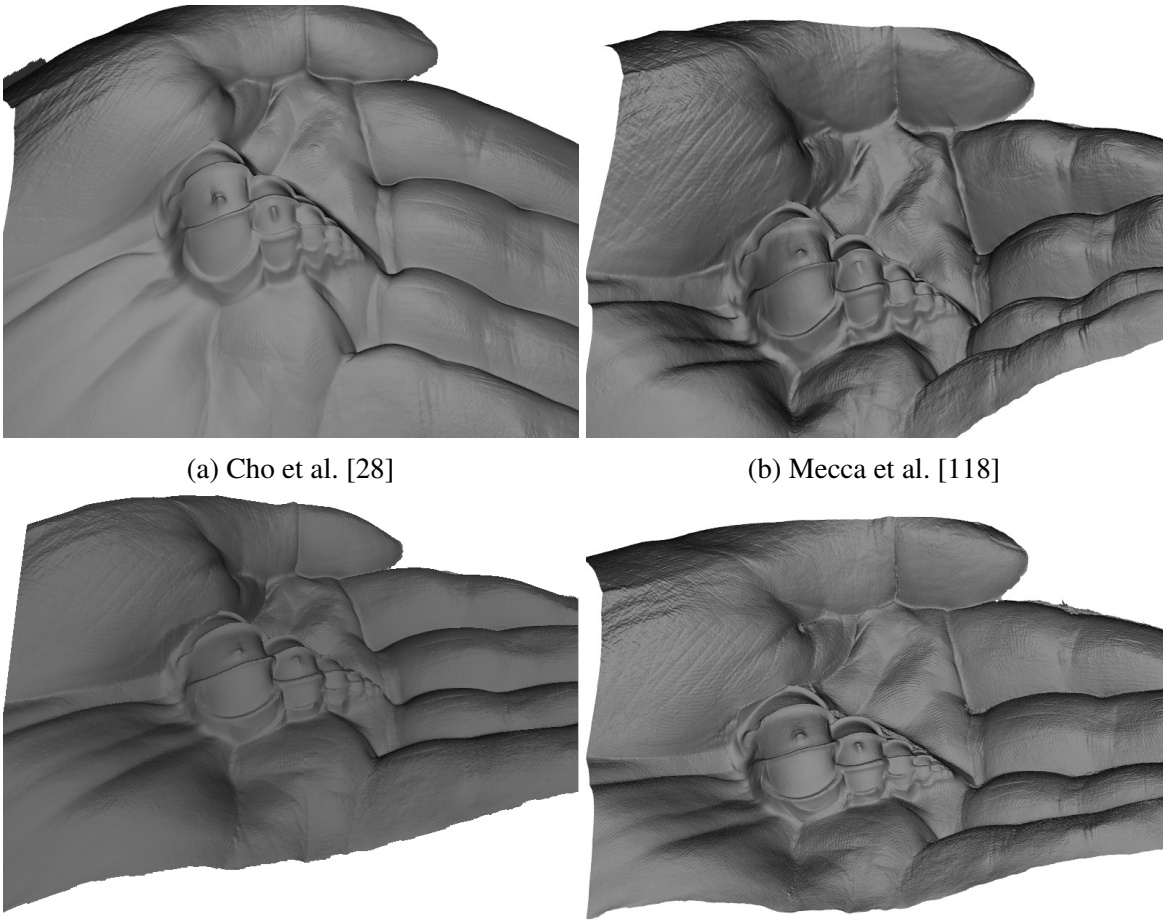
(a) [28] (b) [118] (c) [149] (d) Proposed

Figure 5.9 Qualitative comparison with [28, 118, 149] based on the real datasets of Figure 5.6. [118] is the baseline calibrated method. The comparison demonstrates the importance of the known light source brightness and thus [118] is easily outclassed. The semi-calibrated near field method of [149] results in almost identical quality results as the proposed method.

Finally, zoomed, textured reconstructions of the proposed method are shown in Figure 5.7. Note that the albedo is calculated as raw luminance value and then demosaicing is used in order to retrieve the RGB colours.

5.7 Discussion

In this chapter the semi-calibrated near-field PS problem was presented. The variational approach of the calibrated near field PS problem (explained in Chapter 3) was extended and incorporated into an alternating optimisation framework. The intrinsic brightness of the



(a) Cho et al. [28] (b) Mecca et al. [118]

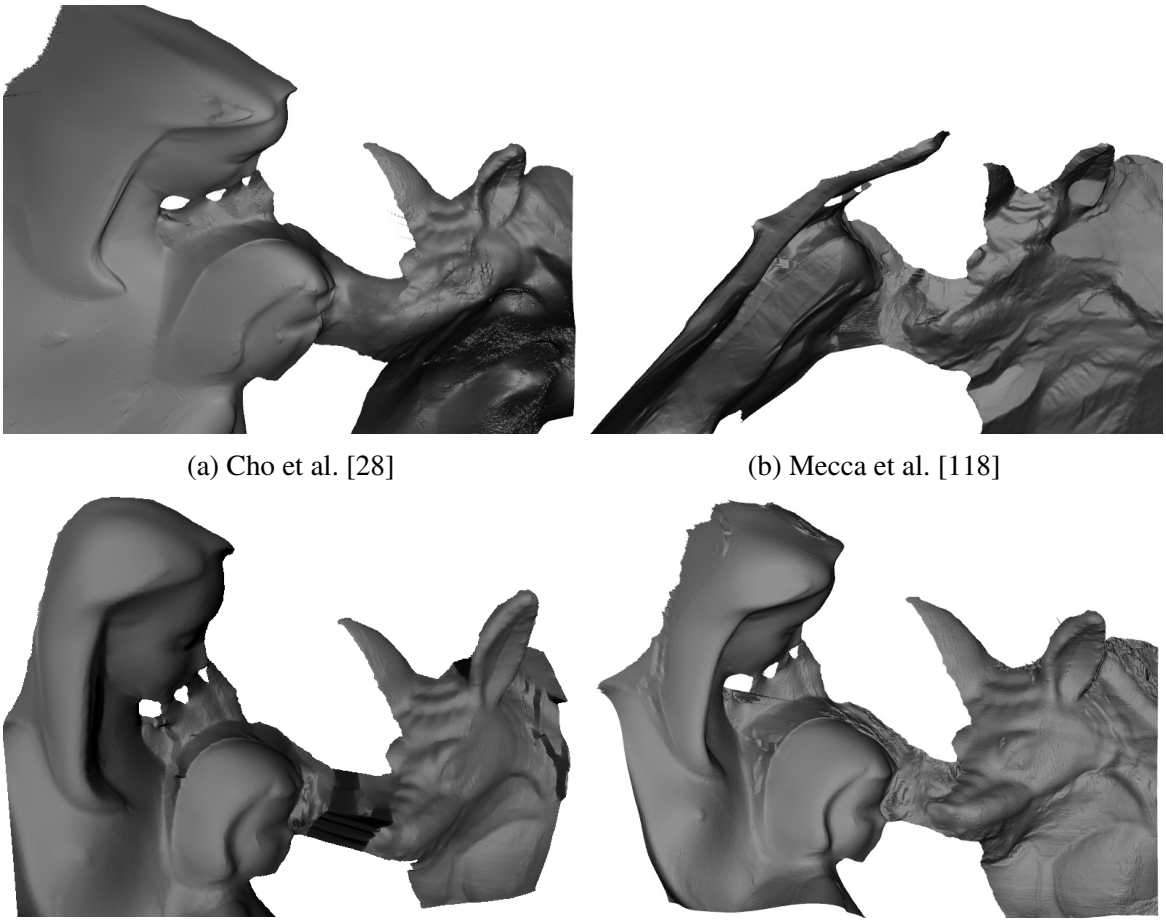
(c) Quéau et al. [149] (d) Proposed

Figure 5.10 Qualitative comparison with [28, 118, 149] on the hand dataset of Figure 5.6. This experiment also confirms the results of Figure 5.9, where [28, 118] are completely outclassed. In addition, the result of [149] also has a small global bending, possibly because it does not handle specular reflection and does not optimise light attenuation.

light sources, as well light attenuation factors, were also optimised, while still retaining the perspective viewing geometry, nearby light-sources and specular reflection modelling.

The proposed approach was evaluated across a number of synthetic and real world datasets containing heterogeneous materials, and obtained state-of-the-art results and even compared favourably with the later method of Quéau et al. [149]. In addition, the ability of the proposed method to estimate intrinsic light source brightness was quantitatively evaluated through LUX meter measurements and all competing methods were outclassed.

Limitations The main limitation of this approach is its inability to deal with surfaces with a vastly different reflectance from the Blinn-Phong specular model or the Lambertian



(a) Cho et al. [28] (b) Mecca et al. [118]

(c) Quéau et al. [149] (d) Proposed

Figure 5.11 Qualitative comparison with [28, 118, 149] using the real dataset of Figure 5.1. This is by far the most challenging dataset as it features discrete objects with varied material properties (marble and plastic), with a depth discontinuity in their boundary, and also shadows and self occlusions. In addition, the dataset was shot with varied exposure so as to push the $\mathbf{\Phi}$ variation/estimation to the limit. As a result [118] fails quite significantly as its result is very deformed. [28] is also performing quite poorly: it avoids significant global deformations as it produces a reasonable estimate of $\mathbf{\Phi}$ (see Figure 5.8b) but still has a number of artefacts e.g. on the child's heads. Finally, [149] is also surpassed (although by a small margin) and this demonstrates the significance of specifically handling specular reflection.

(diffuse) model. In addition, the method has a very high computation cost, mainly because of the shadow map ray-tracing as well as the light attenuation optimisation. Ray-tracing shadows is very expensive due to the depth map parameterisation of the surface, as for each pixel a whole line of pixels needs to be traced to determine visibility. Moreover, this depth map parameterisation also induces bias and unwanted deformations in scenes with multiple discrete objects, which is a generalised problem for most 2D PS methods. This problem

will be dealt with in the next Chapter where the multi-view PS problem is addressed and the proposed solution can reconstruct multiple discrete objects. In addition, the proposed volumetric approach is naturally organised in an octree data structure which significantly speeds up ray-tracing visibility computations.

Future Extensions A straightforward extension of this work would be to replace the point-wise attenuation map estimation with a fully variational optimisation. This could be achieved by stacking the attenuation Equation 5.15 along with the smoothness constraints $a(u, v) = a(u + 1, v)$ and $a(u, v) = a(u, v + 1)$ into a big linear system, which when solved, would generate a more accurate estimate. In the context of semi-calibrated near field PS, an important future task would be to consider non-negligible ambient light (see Chapter 4). This makes the problem much harder to solve since the photometric parallax is reduced by the presence of an additional offset. In addition, image differences do not immediately eliminate the ambient light as this is proportional to the image exposure, which can be assumed unknown in the case examined in this chapter. Finally, the applicability of the proposed light attenuation method estimation in the case of a significantly attenuating medium, such as water, is foreseen.

Chapter 6

Multi-view Photometric Stereo

Figure 6.1 The Multi-View Photometric Stereo Problem: the objective in this chapter is to reconstruct scenes of multiple objects from heterogeneous materials. The images here show three views of a multi-object scene made up of a swede next to a porcelain cup which contains a small tree branch. Readily available SFM software can easily track the camera motion around the scene and create a sparse reconstruction (see Figure 6.2), but creating a dense reconstruction is not trivial.

6.1 Introduction

Up until this point in this thesis, it has been demonstrated how Photometric Stereo has been used to retrieve highly detailed surfaces from a single view. However, obtaining a full 3D object reconstruction, by merging data from different points of view is non-trivial (this will be further explained in Section 6.3.1).

Multi-view stereo, has been the go-to method for obtaining large scale reconstructions. MVS is complementary to PS, as it is mostly reliable for reconstructing specific point features which can be matched in different views and then triangulated (e.g. the eight point algorithm [107]). On the other hand, PS is mostly reliable for continuous and differentiable surfaces and does not perform well around sharp corners and edges. In addition, as seen in Section 2.2.2, PS is more accurate when reconstructing high spatial frequencies rather than low ones.

Multi-View Photometric Stereo (MVPS) approaches have been developed to overcome constraints coming from both sides, in order to deal with: specular highlights Jin et al. [79], Ackermann et al. [3], dynamic scenes Vlasic et al. [183], visibility and occlusions Delaunoy and Prados [39] and mapping of the Photometric Stereo views onto the coarse volume Sabzevari et al. [157], Park et al. [144].

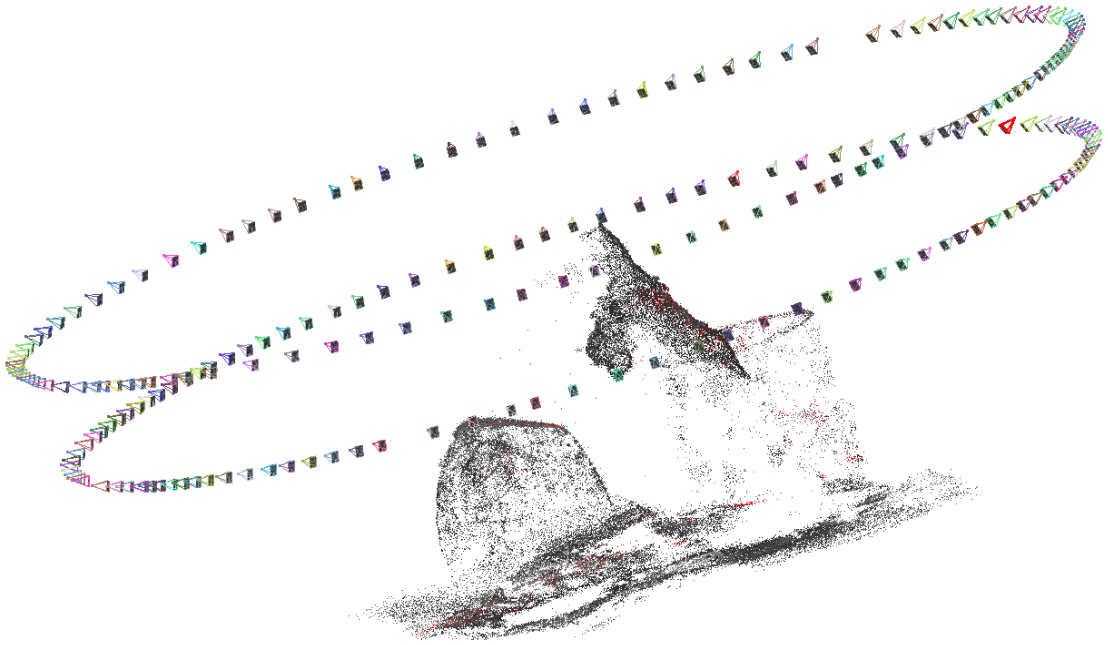


Figure 6.2 Out of the box structure from motion software (VisualSFM [193]) can provide an accurate camera trajectory and a sparse reconstruction around the scene of interest (Figure 6.1) by using a raw video stream as input. The objective of this chapter is to use this information along with images with varied illumination so as to be able to generate an as-dense-as possible reconstruction of the scene.

It is generally accepted that SFM algorithms are reliable for tracking camera motion in a scene and for providing a sparse point cloud reconstruction with most of the research in MVS and MVPS aiming at increasing the accuracy of the reconstruction by generating denser and denser surfaces.

Since implicit parameterisation of volumes has been developed using level-set approaches Malladi et al. [112], Osher and Fedkiw [140], recent advances in parameterising volumes

with signed distance functions (SDFs) [210, 134] have made the multi-view approach prone to be merged with differential formulation of irradiance equation providing shading information [111]. On the other hand, recent PS approaches have moved towards more realistic assumptions, giving consideration to point light sources [120, 149] that make the acquisition process easier by using LEDs in a calibrated setting.

Following the paradigm of other MVPS methods, it is assumed that images from multiple viewpoints and varied illumination (obtained by the capture setup of Figure 1.5) are a given. In addition, it is assumed that the camera trajectory (which also fully calibrates the light source positions as the LEDs are rigidly attached to the camera) is known, and that a low quality reconstruction of the scene also exists.

Contribution This work proposes the following:

- A differential parameterisation of the volume, based on the signed distance function that allows irradiance equation ratios to deal with near-field Photometric Stereo modelling (Section 3.3).
- A variational optimisation that fuses information from multiple viewpoints into a single system. This is an evolution of the 2D Photometric Stereo variational optimisation of Section 3.7.
- An octree implementation capable of retrieving highly accurate volumetric reconstructions in scenes with multiple discrete objects.

6.2 Related works

There are multiple examples in the literature for merging multi-view coarse reconstructions with techniques based on shading information so as to provide high frequency details of the surface [125, 187, 195, 129, 15] rather than topological evolution of the surface Jin et al. [79]. However, regarding the refinement, several methods take inspiration from SFS (Section 2.3) to extract 3D geometry from a single image (MVSfS) and also consider shape refinement coming from single shading cues Wu et al. [197, 198], Barron and Malik [9] and Jin et al. [77, 76]. With the aim to improve the quality of the details and make the reconstruction more robust to outliers, multiple shading images from a single view point are considered. A number of MVPS approaches have been presented [59, 144, 209].

Merging shading information with multi-view images becomes a more complicated problem when considering specular surfaces. Drastic changes in both the shading under

different lighting and the viewing point, modify the appearance of the 3D geometry so that specific approaches have been developed to deal with irradiance equations with a not negligible specular component. Jin et al. [79] exploited a rank constraint on the radiance tensor field of the surface in space, with the aim of fitting the Ward [186] reflectance model. Other approaches reconstructed an unknown object by using a radiance basis inferred from reference objects instead (Treuille et al. [180] and Ackermann et al. [3]). Zhou et al. [209] developed a camera and a hand-held moving light system in order to first capture sparse 3D points and then refine the depth along iso-depth contours Alldrin and Kriegman [5]. A similar handheld system has been developed by Higo et al. [62] where multi-view images were acquired under varying illumination by a handled camera with a single movable LED point light source for reconstructing a static scene.

In order to make the MVPS solvable, additional assumptions have been considered. More specifically, with the aim to compute the camera positions so as to accurately map the Photometric Stereo views, thereby constraining the relative motion of the camera and the object. Hernández et al. [59] captured multi-view images for a moving object under varying illuminations by combining shading and silhouettes assuming circular motion for the computation of the visual hull. Zhang et al. [207] generalised optical flow, Photometric Stereo, multi-view-stereo and structure from motion techniques assuming rigid motion of the object under orthographic viewing geometry and directional lighting. Furthermore, shadows, occlusions or inter-reflections were not considered.

When PS (as well as SfS) has to be integrated with multi-view techniques, the problem of finding the correspondence of pixels with shading information on the 3D surface is crucial. Geometric distortions produced by changes in pose have to be combined with varying illumination. One way to do so is by region tracking considering brightness variations using optical flow [57], parametric models of geometry and illumination [55], or outlier rejection Jin et al. [78]. Okatani and Deguchi [136] proposed a photometric method for estimating the second derivatives of the surface shape of an object when only inaccurate knowledge of the surface reflectance and illumination is given by assuming a probabilistic framework and solved by belief propagation.

Other approaches instead align the shading images with a coarse 3D mesh in order to map the PS data onto the coarse 3D shape [100, 82]. Delaunoy and Prados [39] used a gradient flow approach whereas Sabzevari et al. [157] firstly computed a 3D mesh with structure from motion with a low percentage of missing points and then the mesh was reprojected onto a plane using a mapping scheme [102]. Recently, Park et al. [144] proposed a refinement method by computing an optical displacement map in the same 2D planar domain of the PS images. To do so, they transformed the coarse 3D mesh into parametrised

2D space using a distortion parameterisation technique [165]. The main limitation of these mesh parameterisation techniques is their inability to optimise the object's topology; this is dealt with in Yoshiyasu and Yamazaki [204] where an alternation between mesh and SDF parameterisation is proposed in order to be able to adjust the topology.

In this chapter, a differential approach for MVPS that avoids any mapping procedure is presented. Being inspired by the signed distance function parameterisation and refinement used by Maier et al. [111] for the MVSfS problem, a volumetric parameterisation which handles the differential irradiance equation ratios presented in Chapter 3, is derived. This parameterisation is implemented on an octree (Luebke et al. [110]) which allows to perform fast ray-tracing to estimate shadows and occlusions.

6.3 Signed distance function

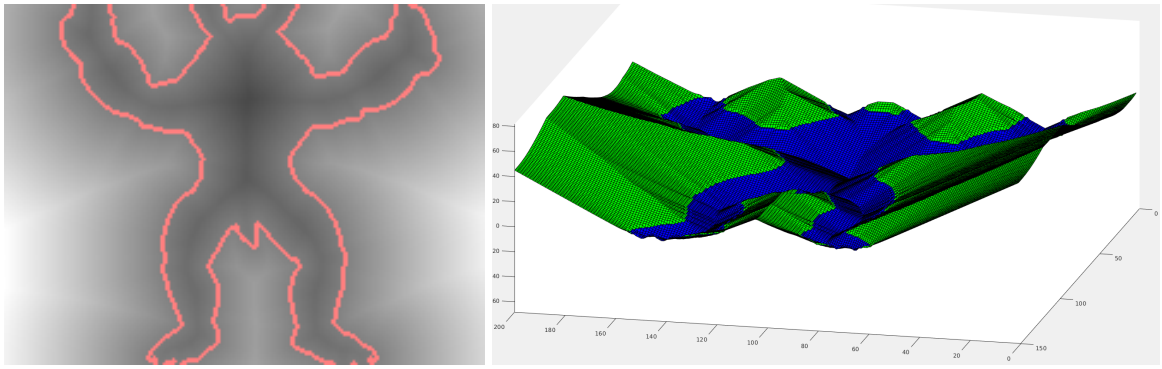


Figure 6.3 Illustration of the Signed Distance Function for the 2D shape shown left (boundary of the "Armadillo"). This can be graphically depicted as an image (*left*), where high grey level values correspond to high distance outside of the shape and low grey level values correspond to high distance inside the shape. Alternatively, the SDF d can be depicted as a surface plot (*right*), i.e. $[x, y, d(x, y)]$. This graphically illustrates that the SDF is continuous everywhere and differentiable "almost everywhere", which is very important for its integration in a differential approach. Of course, the SDF of a 3D shape (i.e. surface) is considered in practice but a 4D plot cannot be easily illustrated.

In order to provide a suitable mathematical characterisation of a collection of solid objects, the implicit surface parameterisation in terms of the *Signed Distance Function*[1] SDF is considered. The SDF is a function $d : \mathbb{R}^3 \rightarrow \mathbb{R}$ that maps every point to its distance to the *closest* surface (Osher and Fedkiw [140]). By convention, points within the object's bounds are assigned negative SDF values, while points outside its bounds are assigned positive

[1] Also called Signed Distance Field.

values. There is an 1 to 1 mapping between the surface and the SDF, and thus the SDF can be used as an implicit surface parameterisation. In fact, the corresponding surface can be extracted from the zero-crossings of the SDF via the marching cubes algorithm [108]. The SDF d has a number of useful mathematical properties. It is continuous everywhere and differentiable "almost everywhere" (see Figure 6.3 for an illustration), except for points which are equidistant from nearby surfaces. Thus, d is suitable for use in a differential approach as the non-differential points could easily be handled as sparse outliers, in a very similar manner to that in which with depth discontinuities in the standard 2D PS framework are handled (see Section 3.7.4). In addition, d satisfies the eikonal equation (see Section 2.2.2):

$$||\nabla d(\boldsymbol{X})|| = 1 \quad \forall \boldsymbol{X} \in \mathbb{R}^3. \tag{6.1}$$

Moreover, for points $\boldsymbol{X}$ in a *vicinity* of the surface (i.e. $||d(\boldsymbol{X})|| \approx 0$), the gradient of d is identical with the respective surface normal N:

$$\boldsymbol{N}(\boldsymbol{X}) = \nabla d(\boldsymbol{X}). \tag{6.2}$$

The property defined in Equation 6.2 is the key to linking the SDF into a PS parameterisation and expressing an irradiance equation in terms of d. This is described in Section 6.4.

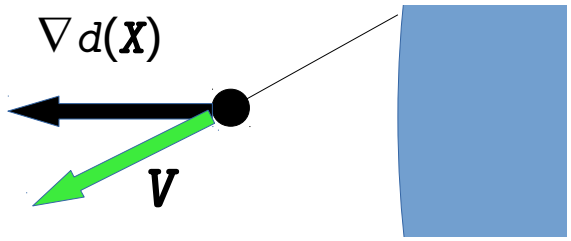


Figure 6.4 The distance of a point infinitesimally close to the surface along a vector $\boldsymbol{V}$ is $\frac{d(\boldsymbol{X})}{\nabla d(\boldsymbol{X})\cdot\boldsymbol{V}}$.

Finally, the following obvious and useful result is noted: for a point $\boldsymbol{X}$ infinitesimally close to the surface, its distance to the surface along a view vector $\boldsymbol{V}$ is $\frac{d(\boldsymbol{X})}{\nabla d(\boldsymbol{X})\cdot\boldsymbol{V}}$ as shown in Figure 6.4.

6.3.1 Naive Photometric Stereo reconstructions merging

Before proceeding to the differential approach for MVPS, the naive solution to the problem is addressed; this solution is the merging of depth maps obtained by using PS from different views. In fact, the SDF has been the traditional way of merging multiple depth maps,

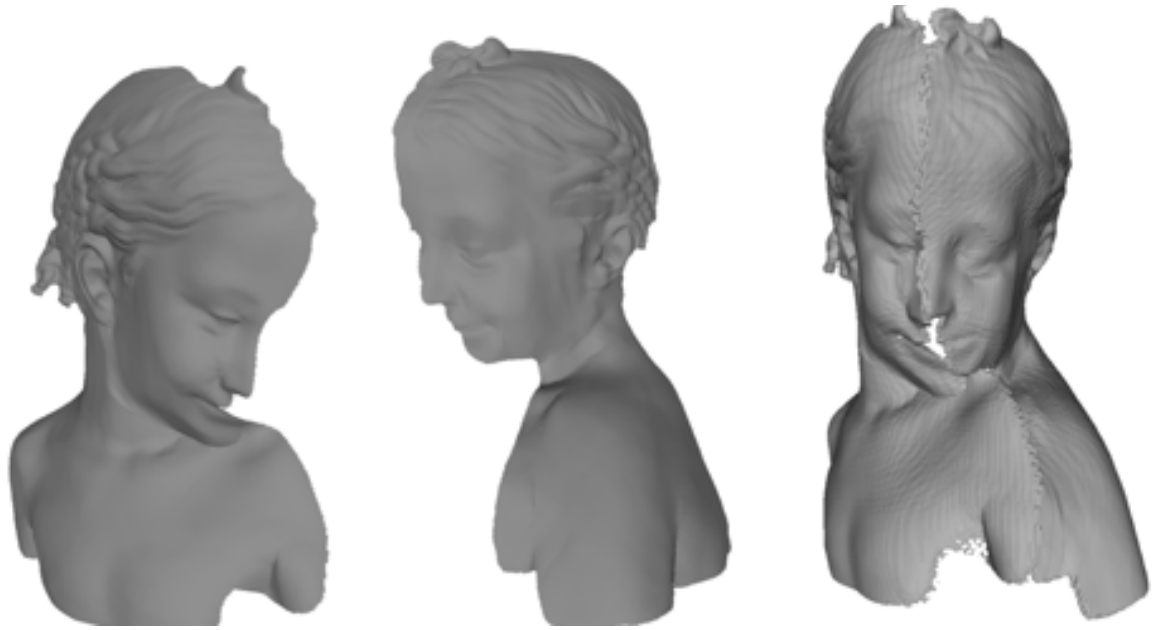
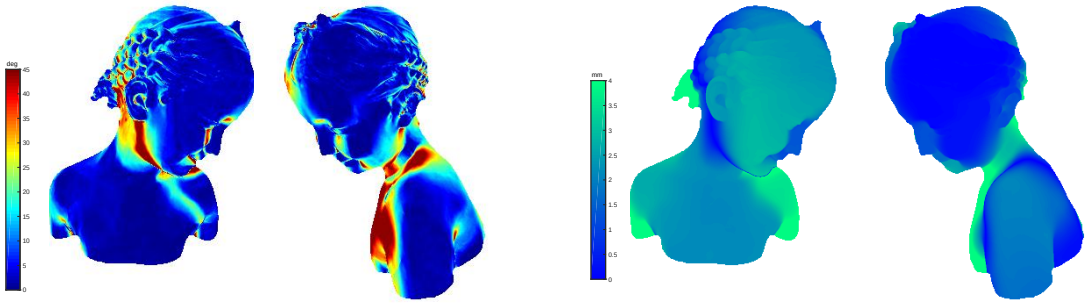


Figure 6.5 Naive PS reconstructions merging. Relatively nice looking reconstructions can be very incompatible and lead to a very bad merging, when using a simple implementation of Curless and Levoy [34]. As shown in Figure 6.6, small normal errors propagate to large depth errors making the different pieces merge incompatibly and lead to severe artefacts in the boundary region.

introduced in Curless and Levoy [34]. Various SDF-based fusion algorithms are still very popular such as Kinect fusion [130] as well as more recent methods such as Maier et al. [111]. However, the problem with PS reconstructions is that small, sparse errors in normals (which essentially provide all the information) lead to systematic deformations in the depth maps as shown in Figure 6.6.

In addition, the use of a robust merging algorithm is probably not going to solve the incompatibility problems shown above. For example Ylimäki et al. [203] employed median filtering and used estimates of the uncertainty of depth to calculate a covariance matrix for the calculation of the Mahalanobis distance of the 3D clouds being merged. The systematic depth errors in the PS depth maps, mainly arising from smoothing out depth discontinuities, are not sparse (see Section 2.2.1) and thus these approaches are expected to underperform.

Finally, one main reason for the success of Kinect fusion-like algorithms is the ability of structured light sensors to provide a large amount of depth data, typically at least 30 frames per second. In fact, Curless and Levoy [34] showed that their merging of depth maps converges to the true SDF assuming a large number of views. However, obtaining a large number of PS reconstructions is impractical, mainly for the following two reasons. Firstly, PS acquisition requires multiple images per view and thus it is likely to require some non-negligible capture



(a) Normal MAE: 2.6^o & 4.3^o . (b) Depth RMS Errors: 2.5mm & 1.5mm

Figure 6.6 Demonstration of how small, sparse normal map errors lead to severe depth map errors. Note that 4mm (light green) is 20% of the diameter of the 20mm object.

time. A notable exception is Brostow et al. [21], where RGB channels are exploited in order to capture three images at the same time and perform real time reconstruction. However, this approach is critically dependent on the assumption of equal albedo on all three channels. In addition, the approaches presented in Chapters 3 to 5 require more than three images per view (especially if specularity coefficients and light source brightness are calculated) and thus a multi-spectral camera would be required. Finally, the variational optimisation discussed in Section 3.7 is computationally expensive and not running in real time. The approach presented here overcomes all of these limitations. A unified PDE is formed on the SDF and a variational optimisation, merging information from all views at the same time, is presented. The approach does not require too many views, the requirement being just a sufficient coverage of the object's surface.

6.4 Differential multi-view Photometric Stereo

Similarly to [198, 210, 111] that used the SDF for single image shading refinement, the SDF is considered for the irradiance equation to derive a differential multi-view Photometric Stereo formulation. It is assumed to have n_{ps} images (i.e. light sources) for each known camera configuration C_q (that is $n_{ps}(C_q j)$, $j = 1, \dots n_{views}$). The matrix C_q is used to contain the camera configuration as $C_q = [R_q, \boldsymbol{T}_q]$, where R_q and $\boldsymbol{T}_q$ are rotation matrix and translation vector, linking image plane point $(u, v)_q$ to world 3D point $\boldsymbol{X}$ as:

$$(u, v)_q = \Pi(R_q \boldsymbol{X} + \boldsymbol{T}_q) \tag{6.3}$$

where Π denotes perspective projection as defined in Equation 3.6.

Following the same paradigm as the rest of this thesis, to exploit the monocular aspect of the PS problem, image ratios for the Lambertian shading model (Section 2.1.1) are considered, assuming calibrated nearby LED light sources. Thus, the respective irradiance equation is:

$$i_j(u, v) = i_j(\boldsymbol{X}) = \rho(\boldsymbol{X}) a_j(\boldsymbol{X}) \overline{\boldsymbol{N}}(\boldsymbol{X}) \cdot \overline{\boldsymbol{L}}_j(\boldsymbol{X}) \tag{6.4}$$

where $(u, v) \in \mathbb{R}^2$ is the image-plane projection of the 3D point $\boldsymbol{X}$ as defined in Equation 6.3.

It is very important to emphasise that this irradiance equation is considered for points $\boldsymbol{X}$ on the 3D space in contrast to the rest of the thesis which considers irradiance equations for 2D image plane points. Thus $i_j(\boldsymbol{X})$ is the reflected radiance which can be calculated by projecting that point $\boldsymbol{X}$ onto the image plane and thus $i_j(u, v) = i_j(\boldsymbol{X})$. In practice this projection will require a linear interpolation as discussed later in Figure 6.8 and it is only valid if the 3D point $\boldsymbol{X}$ is not occluded.

In addition, the volumetric irradiance equation 6.4 is only valid for points in the vicinity of the surface, as $\rho(\boldsymbol{X})$ and $\boldsymbol{N}(\boldsymbol{X})$ are only meaningfully defined for surface points. As pretty much the entirety of PS literature assumes continuity, it is assumed that properties that are exactly true for surface points, will be approximately true for points close to the surface.

In contrast, the light attenuation $a_j(\boldsymbol{X})$ and the lighting vector $\boldsymbol{L}_j(\boldsymbol{X})$ can be defined for every 3D point $\boldsymbol{X}$ using the definitions presented in Section 3.3, regardless of the existence of a surface or the value of the SDF around that point. Again assuming no occlusions, a_j and $\boldsymbol{L}_j$ are trivially computed with the respective formulas (assuming calibrated setup).

6.4.1 Modelling with image ratios

To simplify the non-linear normalisation of the normal as well as the albedo, the ratio method (Section 3.2) is followed. Indeed, dividing equations for images j and k as in Equation 6.4 (ignoring dependence of all variables in $\boldsymbol{X}$ for clarity), gives:

$$\frac{i_j}{i_k} = \frac{a_j \boldsymbol{N} \cdot \overline{\boldsymbol{L}}_j}{a_k \boldsymbol{N} \cdot \overline{\boldsymbol{L}}_k} \quad \Rightarrow \quad \boldsymbol{N} \cdot (i_j a_k \overline{\boldsymbol{L}}_k - i_k a_j \overline{\boldsymbol{L}}_j) = 0. \tag{6.5}$$

It is noted that images j and k do not necessarily need to be from the same point of view C_q but this minimises the inaccuracy caused by not being exactly on the surface boundary (it has been assumed that points where the irradiance equation is considered are near the surface but of course if the surface was known precisely, the reconstruction problem would have already been solved). Essentially this projection matches image points $i_j(u_j, v_j)$, $i_k(u_k, v_k)$ which are guaranteed to be projections of the same 3D point $\boldsymbol{X}$ only for images from the same point of view.

By substituting the parameterisation of the normal from Equation 6.2, the following albedo independent, homogeneous linear PDE is obtained:

$$\boldsymbol{B}_{jk}(\boldsymbol{X}) \cdot \nabla d(\boldsymbol{X}) = 0 \;\text{ where }\; \boldsymbol{B}_{jk} = i_k a_j \overline{\boldsymbol{L}}_j - i_j a_k \overline{\boldsymbol{L}}_k. \tag{6.6}$$

The geometrical meaning of Equation 6.6 is the extension of the 3D volumetric reconstruction of the PDE approach presented in Mecca et al. [121]. In fact, the PS model still consists of a homogeneous linear PDE where the tangentiality of $\boldsymbol{B}_{jk}$ on the surface is by definition the zeroth level set of the SDF.

There are two important differences to [121]. First of all, $\boldsymbol{B}_{jk}(\boldsymbol{X})$ does not directly depend on d, due to the fact that the relevant quantities are expressed on a global coordinate system independent of the existence of a surface. Thus the equation is linear (instead of quasi-linear) and $\boldsymbol{B}$ is easily calculated from image data. Of course, there is the indirect dependence on the fact that Equations 6.4 to 6.6 are only true in the vicinity of the surface. In fact, the correct way of thinking about the 3D irradiance equations is *If there is a surface near a point there, its orientation satisfies that equation*. The other difference of Equation 6.6 to the unified equation of 2D PS (3.18) is that it is homogeneous and thus ill-posed. Directly solving for d is impossible, even if d is known on a number of points. The solution to this problem is to incorporate a Tikhonov regulariser (see below).

An interesting observation is that Equation 6.6 is conceptually similar to the iso-depth curves in the work of Zhou et al. [209]. Nonetheless, the SDF formulation is a more "natural object-centered" depth and this allows for a unified optimisation as described in the next section. In order to simplify the notation, the pair jk is renamed as p.

6.4.2 Multi-view Photometric Stereo as a weighted least square problem

In order to consider PS images coming from different views into a single mathematical framework, the following weighted version of Equation 6.6

$$w_q(C_q, \boldsymbol{X})\boldsymbol{B}_q(\boldsymbol{X}) \cdot \nabla d(\boldsymbol{X}) = 0 \tag{6.7}$$

is stacked in a single system to get:

$$\begin{bmatrix} [w_1(C_1, \boldsymbol{X})\boldsymbol{B}_1(\boldsymbol{X})]^\intercal \\ [w_2(C_2, \boldsymbol{X})\boldsymbol{B}_2(\boldsymbol{X})]^\intercal \\ \vdots \end{bmatrix} \nabla d(\boldsymbol{X}) = \boldsymbol{0} \tag{6.8}$$

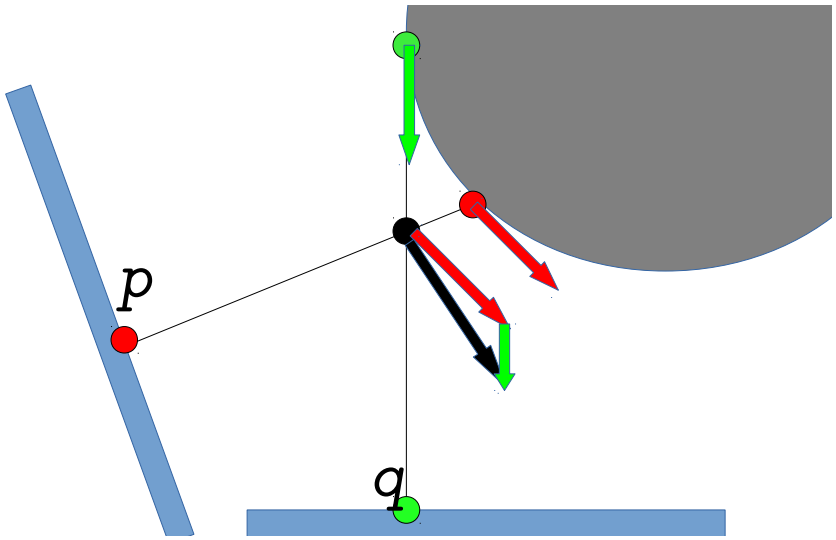


Figure 6.7 Considering photometric ratios from the same point of view reveals the surface orientation (or equivalently the SDF levelset) in the projection of that point to the surface along the viewing ray. In the figure, one pair from image p reveals the SDF levelset orientation at the red point, and another from image q at the green one. Then the actual SDF levelset (black arrow) should be a weighted sum of these orientations, with weights inversely proportional to the distance to the surface between these viewing rays which equals $\frac{d(\boldsymbol{X})}{\nabla d(\boldsymbol{X})\cdot \boldsymbol{V}(\boldsymbol{X})_p}$ and $\frac{d(\boldsymbol{X})}{\nabla D(\boldsymbol{X})\cdot \boldsymbol{V}(\boldsymbol{X})_q}$ respectively (see Figure 6.4). Hence, it is reasonable to set weights to $\nabla d \cdot \boldsymbol{V}_p = \boldsymbol{N} \cdot \boldsymbol{V}_p$ and $\boldsymbol{N} \cdot \boldsymbol{V}_q$ respectively. Of course all of this discussion assumes continuity and that the point in question (black) is close to the actual surface.

$w_q(C_q, \boldsymbol{X}) = \max(\boldsymbol{N}(\boldsymbol{X}) \cdot \boldsymbol{V}_q(\boldsymbol{X}), 0)$ and $\boldsymbol{V}_q(\boldsymbol{X})$ denotes the viewing vector on the volume for the camera position C_q. This weight term w_q is essentially a measure of visibility aimed at increasing the significance for equations corresponding to views that have a direction viewing angle at a particular point in space. For an additional explanation of the geometrical meaning of this weighted sum, see Figure 6.7. In addition, in case of a self-occlusion, the corresponding weight is set to 0. The addition of this weight term does not change the form of the equation which is still a homogeneous linear PDE.

The resulting system then counts $\sum_{q=1}^{N_{views}} \binom{N_{ps}(C_q)}{2}$ equations as (6.8), and with the aim to solve it as a least square problem, the normal equations are considered

$$B(\boldsymbol{X})\nabla d(\boldsymbol{X}) = 0 \tag{6.9}$$

with:

$$B = [w_1(C_1, \boldsymbol{X})\boldsymbol{B}_1(\boldsymbol{X}), w_2(C_2, \boldsymbol{X})\boldsymbol{B}_2(\boldsymbol{X}), \ldots] \begin{bmatrix} [w_1(C_1, \boldsymbol{X})\boldsymbol{B}_1(\boldsymbol{X})]^\intercal \\ [w_2(C_2, \boldsymbol{X})\boldsymbol{B}_2(\boldsymbol{X})]^\intercal \\ \vdots \end{bmatrix}$$

being a positive, semi-definite, 3×3 matrix.

Rank correction The geometrical constraint coming from (6.6) ensures that all the vector fields $\boldsymbol{B}_q(\boldsymbol{X}) \in \mathbb{R}^3$ span the same bi-dimensional space $\forall \boldsymbol{X}$ of the volume as they define the level-set of the SDF. This means that under ideal circumstances, the rank of B in Equation 6.9 should be exactly 2. However, due to numerical approximations this is never exactly true; this constraint can been enforced by using the eigenvalue decomposition of B hence:

$$B = Q\Lambda Q^\top = Q \begin{bmatrix} \Lambda_1 & 0 & 0 \\ 0 & \Lambda_2 & 0 \\ 0 & 0 & \Lambda_3 \end{bmatrix} Q^\top \text{ with } \Lambda_1 \geq \Lambda_2 \geq \Lambda_3 \text{ and setting } \Lambda_3 = 0. \quad (6.10)$$

Another reason why B cannot be full rank is the Eikonal Equation 6.1. In fact, if $B\nabla d = 0$ and B is full rank then $\nabla d = 0$ which contradicts $\|\nabla d\| = 1$. In practice this would mean that d does not closely resemble an implicit surface parameteristation and that extracting a surface with the marching cubes algorithm would produce unpredictable results. Thus this rank correction step increases numerical stability.

6.5 Variational resolution

In this section, the variational solver introduced in Section 3.7 is adjusted in order to be able to deal with the novel 3D-SDF parameterisation introduced in the previous section. First of all, as already discussed, Equation 6.9 is rank deficient, and thus a closed form computation of the volume is not possible. This is not surprising as the PS constraint encodes orientation and should not convey any information about the rate of growth of the SDF; this information is encoded in the Eikonal Equation 6.1. Indeed, integrating Equation 6.1 with Equation 6.9 would lead to a much more constrained problem which would however be much harder to solve as Equation 6.1 is non-linear. Finally it is noted that there is a single parameter ambiguity due to the fact that the equation depends on ∇d and not d directly (which is very similar to the problem presented in Section 3.7.3). The actual reason for this ambiguity is the convention of representing the surface as the zeroth level-set of the SDF; the whole differential parameterisation described above would be consistent with any level set value.

All of these ambiguity reasons are resolved by adopting a Tikhonov regulariser of the form $d(\boldsymbol{X}) = d_0(\boldsymbol{X})$ where $d_0(\boldsymbol{X})$ is some initial estimate of the SDF; this is calculated as the distance transform of an initial surface estimate. This estimate can be obtained with MVS or any other 3D reconstruction method and as later demonstrated in the experimental Section 6.7, it is not required to be very accurate. In fact, as it has been assumed that camera positions

and orientations are known, it is not unreasonable to expect to have an initial estimate of the geometry as well, as a result of the output of standard SFM pipelines (e.g. VisualSFM [193]). Thus the regularised version of the problem becomes:

$$\min_{d}(\|B(\boldsymbol{X}) \cdot \nabla d(\boldsymbol{X})\|_{\ell_2} + \lambda \|d(\boldsymbol{X}) - d_0\|_{\ell_2}). \tag{6.11}$$

The use of ℓ_2 could in theory be replaced with ℓ_1 as described in Section 3.7.4, however a choice was made not to do that for the following reasons. Firstly, enforcing the rank 2 constraint is much easier after considering the ℓ_2 solution (Equation 6.10), as the resulting matrix is 3x3; otherwise, performing SVD on a $N_{views}\binom{N_{ps}}{2} \times 3$ matrix per point on the domain (see next section about the discretisation) would be required. This would therefore be orders of magnitude more computationally expensive. In addition, the use of the ℓ_2 vastly reduces the memory requirements as discussed in Section 3.7.2.

6.5.1 Discretisation

In this section, how the PDE of Equation 6.11 can be discretised and solved over a finite set of 3D points is described. This set $\Omega \subset \mathbb{R}^3$ of 3D points $\boldsymbol{X} = (x, y, z)$ need to be close to the surface, as this is required if the normal parameterisation of Equation 6.2 is to be valid. In addition, it is convenient to think of these points as voxels with a finite size s, as this is crucial for projecting to images and hence determining $i(x, y, z)$ as illustrated in Figure 6.8.

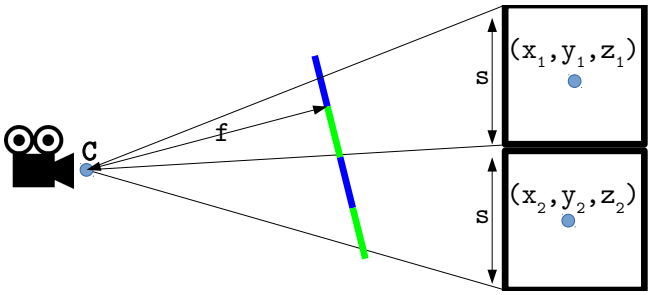


Figure 6.8 Discretising the volume into a set of voxels with finite size s is imperative for determining correspondence to pixels in order to perform linear interpolation and thus determine $i(x, y, z)$. This is shown by colouring successive pixels on the image plane with different colours, in order to emphasise the possibility of 1 to many mapping between voxels and pixels. Note that C is the camera center and f the focal length of the camera.

After having established a connected set of voxels, the next step involves expressing the partial derivatives using finite differences. This is done in a very similar fashion as in the 2D case discussed in Section 3.7.1. Denoting as $\boldsymbol{D}$ the vector stacking all $d(\boldsymbol{X})$ for all $\boldsymbol{X} \in \Omega$,

one obtains:

$$\frac{\partial \boldsymbol{D}}{\partial x} = G_x \boldsymbol{D} \ , \ \frac{\partial \boldsymbol{D}}{\partial y} = G_y \boldsymbol{D} \ , \ \frac{\partial \boldsymbol{D}}{\partial z} = G_z \boldsymbol{D} \tag{6.12}$$

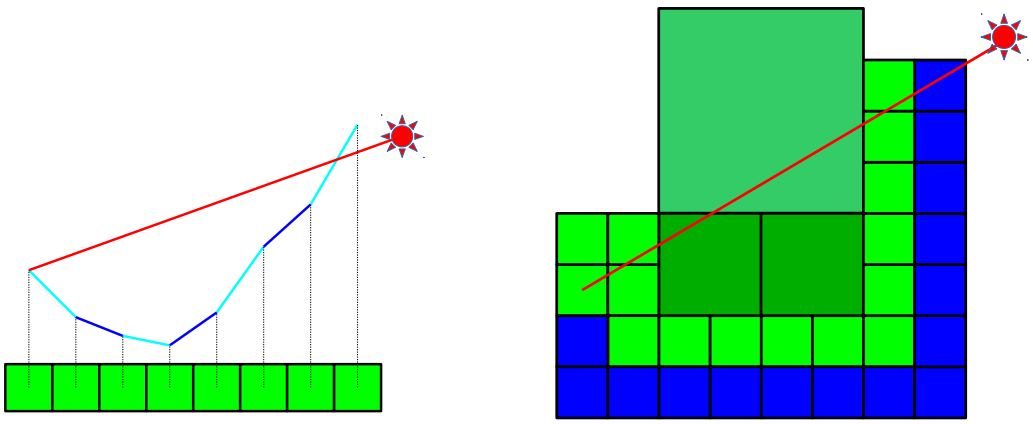
where G_x, G_y and G_z are sparse matrices with -1 and 1 on the non-zero entries. Equation 6.12 is a direct 3D generalisation of Equation 3.29. Thus, Equation 6.11 is discretised in the form:

$$A_s \boldsymbol{D} = \boldsymbol{B}_s \tag{6.13}$$

where A_s is symmetric positive definite and thus the system can be solved with Cholesky factorisation or conjugate gradients in the case of a very large domain.

6.5.2 Octree implementation

To manage the required set of voxels $\in \Omega$ described above, an octree structure is used. Ω is defined at the leafs of the tree, and voxel neighbours, which are essential for computing finite differences, are found by bottom up traversal of the tree. The use of the octree structure greatly reduces the memory requirements for the representation of the SDF of the surface, and most importantly, greatly accelerates visibility computations, shown in Figure 6.9 and explained below.



(a) Depth map (b) Octree

Figure 6.9 Illustration of how the proposed octree SDF parameterisation accelerates ray-tracing used for visibility estimations. In (b), shades of green denote voxels with positive d and shades of blue denote those with negative d. An occlusion is detected when the tracing of the red ray moves from a green to a blue voxel and thus it means that it has intersected the surface. As the size of the voxels is variable, fewer of them need to be examined in contrast to (a), where the ray is traced on a fixed size depth map.

Visibility Estimation In order to deal with scenes with a complex geometry and potentially, with multiple objects, occlusions need to be addressed. This is performed by ray-tracing lines from each voxel to each light source and camera, by using the current estimate of geometry to check for cast shadows and occlusions. Whenever an occlusion/shadow is detected, the relevant weight in Equation 6.7 is set to 0.

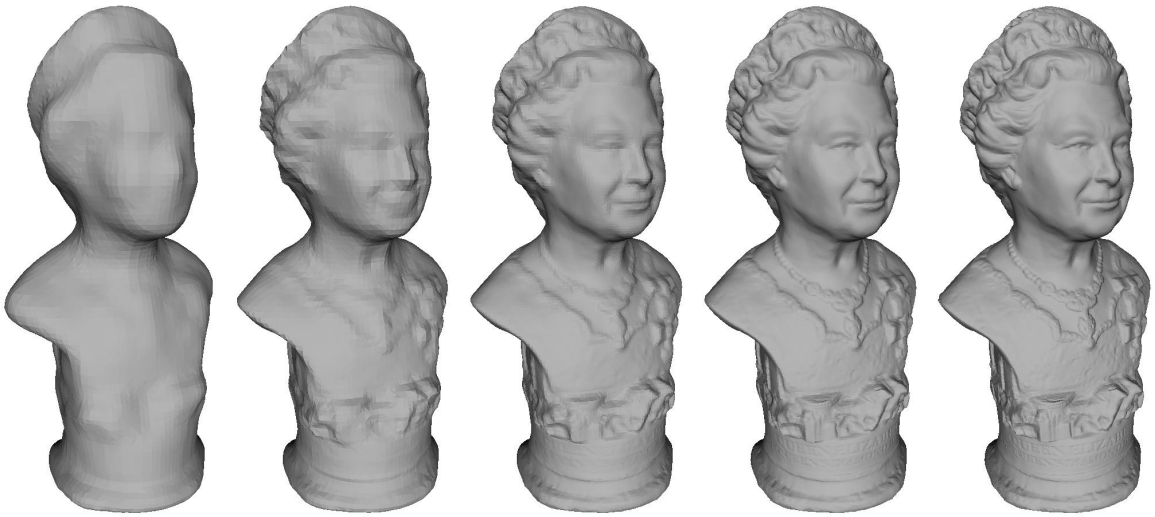
Ray-tracing complexity Octrees have been known to the graphics community for a long time (e.g. Luebke et al. [110]) as they offer better theoretical algorithmic complexity for various tasks such as visibility calculations and collision detection compared to depth maps and triangle meshes. More specifically, for the visibility computation task, assuming a voxel grid of dimension $n \times n \times n$, then the number of voxels around the surface, which are going to be the leaf nodes of the tree where all the calculations are performed, are expected to be $O(n^2)$. Then, for each voxel, visibility calculations only require examining $O(\log n)$ (as illustrated in Figure 6.9b) as opposed to $O(n)$ operations on a depth map parameterisation. The complexity is even worse for a triangle mesh parameterisation (e.g. Park et al. [144]) where there is no notion of spatial ordering (i.e. any 2 triangles could be nearby in space) and thus all $O(n^2)$ pairs of ray-traingle intersections need to be examined. In practice, the logarithmic complexity means that for the datasets examined in Sections 6.7, the ray-tracing operation only required a few seconds, on a several million voxels grid.

6.5.3 Incremental surface reconstruction

Up until this point it has been discussed how the PDE 6.11 can be discretised into a linear system (Equation 6.13) which is solved on the domain of the leafs of an octree. This section further elaborates on the whole iterative procedure.

First of all, it must be emphasised that there is a need for an iterative procedure and a high quality geometry estimate cannot be computed in a single step. This is so because of the following: firstly, the PDE for d (Equation 6.11) implicitly depends on d as this is used for computing visibility with the ray-tracing procedure explained above. Most importantly, the validity of the whole differential approach crucially depends on $\boldsymbol{N}(\boldsymbol{X}) = \nabla d(\boldsymbol{X})$ which is only true for points $\boldsymbol{X}$ close to the surface, which of course is unknown.

Therefore the following iterative procedure is performed: firstly an octree is initialised to densely pack around the surface. The tree is set to a low depth so that the voxels are "relatively big" and thus their centres are relatively close to the surface (i.e. the absolute value of d is comparable to the voxel size s), see Figure 6.10a. Then, Equation 6.11 is solved on the leafs of the tree and thus d values are updated (see Figure 6.10b). Finally, leafs where the absolute value of SDF is smaller than 2 voxel sizes are subdivided and thus the procedure



(a) 1.38mm (b) 1.38mm (c) 0.69mm (d) 0.35mm (e) 0.17mm

Figure 6.10 Incremental refinement of the geometry from the initial estimate (*left*) to the final estimate(*right*). At each step, the voxel size is halved, thus going from 1.38mm to 0.69mm to 0.35mm to 0.17mm. Note that it is not necessary to perform marching cubes after each iteration; here it is only done for illustration purposes.

repeats and approximates the geometry even more. The terminating condition is when the voxels are small enough so that their projection on the image planes is smaller than the pixel size (see Figure 6.8) and thus the maximum obtainable resolution has been reached. When this happens, the reconstructed surface is computed with the Marching cubes variant of [85]. The whole procedure is illustrated in Figure 6.10 and summarised in Algorithm 4. For the example of Figure 6.10, the octree was initialised with depth 7 and after 3 subdivision steps, reached depth 10, reducing the voxel size by a factor of 8 and increasing the voxel count from around 96K to around 4.2M. The resulting final surface consists of 1.06M vertices and 2.1M triangles.

6.6 World scale estimation

In order to parameterise light propagation and attenuation in the near field, the calibrated PS model of Equation 6.6 requires an initial estimate (i.e. d_0 in Equation 6.11) with an approximately correct size. In addition, camera motion in the scene is required to be known in absolute numbers. Note that the camera motion fully determines the light source positions as well as they are rigidly attached to the camera. However, any monocular SFM algorithm suffers from scale ambiguity, meaning that camera translations as well as initial surface estimates are correct up to an unknown scale factor i.e. $\boldsymbol{X} = \boldsymbol{X}_0\xi$ with ξ the unknown scale

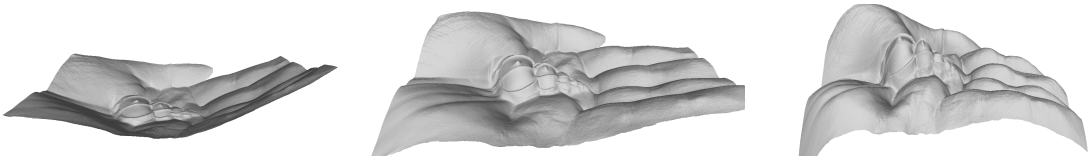
Algorithm 4: Multi-View Photometric Stereo

Input: Images, calibrated light sources & camera trajectory, initial geometry as $\{\boldsymbol{X}_0\}$
Output: Triangle mesh output
Initialisation:
world scale estimation § 6.6;
Poisson reconstruction[84] & Distance transform to get $d_0(\boldsymbol{X})$;
Octree initialisation, leaf size s;
while *Voxel projections* > 1 *pixel (Fig 6.8)* **do**
  set Ω at tree leaves;
  raytrace visibility (Fig 6.9) ;
  calculate $i_j(\boldsymbol{X})\ \forall j\ \&\ \forall \boldsymbol{X} \in \Omega$ (Fig 6.8) ;
  calculate fields $a_j(\boldsymbol{X}),\ \boldsymbol{L}_j(\boldsymbol{X})\ \forall j\ \&\ \forall \boldsymbol{X} \in \Omega$ (Eq 3.13);
  calculate field $\boldsymbol{B}_{jk}(\boldsymbol{X})\ \forall (j,k)\ \&\ \forall \boldsymbol{X} \in \Omega$ (Eq 6.6);
  compute $B(\boldsymbol{X})$ and set $\operatorname{rank}\big(B(\boldsymbol{X})\big) = 2\ \forall \boldsymbol{X} \in \Omega$ (Eq 6.10);
  compute system (Eq 6.13) and solve with conjugate gradients to get $d(\boldsymbol{X})\ \forall \boldsymbol{X} \in \Omega$;
  set $\Omega^{new} : |d(\boldsymbol{X})| < 2s$;
  subdivide voxels $\in \Omega^{new}$, hence $s \leftarrow 0.5s$
end
Marching cubes [85] ;

and $\boldsymbol{X}_0$ the initial estimate. This is not really surprising. As already discussed in Section 3.7, the calibrated near field PS model (as well as the semi-calibrated model of Chapter 5), also requires an initialisation of the correct scale, usually done as a plane at the mean depth of the scene. Having an initialisation of a wrong scale not only severely degrades the reconstruction quality, but also increases the image re-projection error (the error obtained if the computed geometry is substituted into the irradiance equation) as noticed by Quéau et al. [149]. This effect is easily verified experimentally as shown in Figure 6.11.

The dependence of the final re-projection error on the world scale has some value for the automatic estimation of the mean distance for the standard 2D PS problem, although the practical performance is to be experimentally verified, as the example in Figure 6.11 shows only a very small error variation. However, in the context of MVPS this observation is very valuable for calculating the scale of the problem. Indeed, as different scales cause different non-linear deformations on the geometry, it is not unreasonable to assert that only the correct scale will make a reconstruction consistent with the initial MV estimate $\boldsymbol{X}_0$, and that this consistency could be quantitative measured through the re-projection error.

More precisely, given a set of oriented 3D points $\boldsymbol{X}_0$ with normals $\boldsymbol{N}(\boldsymbol{X}_0)$ and a scalar ξ, the scaled points become $\boldsymbol{X} = \boldsymbol{X}_0\xi$. Then, $L(\boldsymbol{X}) = \boldsymbol{X} - \boldsymbol{P}_j = \boldsymbol{X}_0\xi - \boldsymbol{P}_j$ becomes a function of ξ and also a using Equation 3.13. Then, ρ can be computed using least squares on $i = \rho a(\xi)\boldsymbol{N} \cdot \boldsymbol{L}(\xi)$, so ρ is implicitly a function of ξ. On the contrary, $\boldsymbol{N}$ is not dependent on



(a) 9cm: Err 0.01336 (b) 18cm: Err 0.0129 (c) 36cm: Err 0.0139

Figure 6.11 Demonstration of how the initial mean distance estimate affects the final reconstruction of standard near-field PS (Chapter 5) on the hand dataset of Figure 5.6. As can be seen, a too small mean distance flattens the reconstruction; too large a distance makes it non-linearly stretch. This in turn affects the final image re-projection error which can be used as a metric in order to automatically estimate the correct mean distance. In addition, as the deformation is non-linear, only the correct scale would be consistent with estimates obtained using other methods, e.g MVS.

ξ. Thus, the overall function to be minimised is:

$$E(\xi) = \sum_{\boldsymbol{X}_0, j} \| i(\boldsymbol{X}_0 \xi)_j - \rho(\boldsymbol{X}_0, \xi) a(\xi)_j \boldsymbol{N} \cdot \boldsymbol{L}_j(\xi) \| \tag{6.14}$$

where the sum is performed over all points $\boldsymbol{X}_0$ and all images j (in all views). $E(\xi)$ is highly non-linear and non-convex but can be minimised using non-linear simplex[2] by starting from a reasonable estimate obtained with a ruler measurement. This is shown in Figure 6.12.

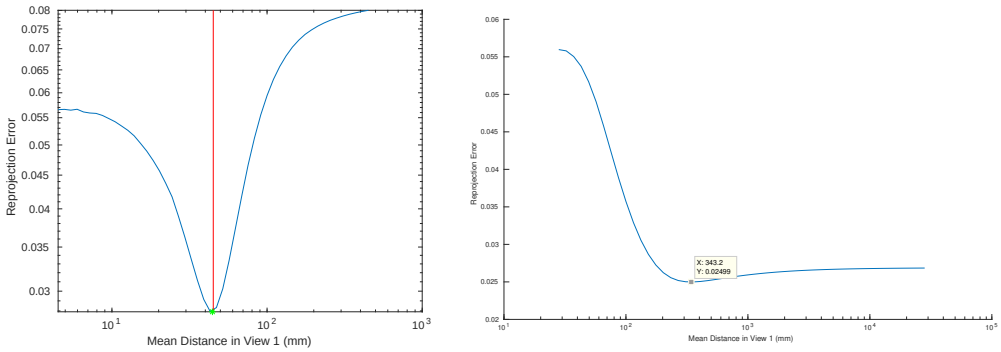


Figure 6.12 Minimisation of Equation 6.14 for estimating the world scale for synthetic data (*left*) and real data (*right*). Note that as ξ is not very meaningful, the scaling on the horizontal axis is the mean distance from the first view to the object. The synthetic experiment is performed on the ground truth mesh of the Armadillo and the obtained minimum (green point) equals, within machine precision, the ground truth (shown in red). The more interesting experiment is the real one performed using the mesh of Figure 6.2, where indeed the error function has a clear minimum.

[2]Using Matlab's *fminsearch* with default parameters.

6.7 Experiments

The proposed algorithm was implemented in Matlab (using $\lambda = 0.05$ for the regulariser weight). The reconstruction time using non-optimised Matlab code, was 15-20 minutes on a i7-4960X CPU with a peak memory consumption of 20-25GB. SDF was computed, (depending on dataset), in around 4-6 M voxels of an approximate size of 0.2mm.

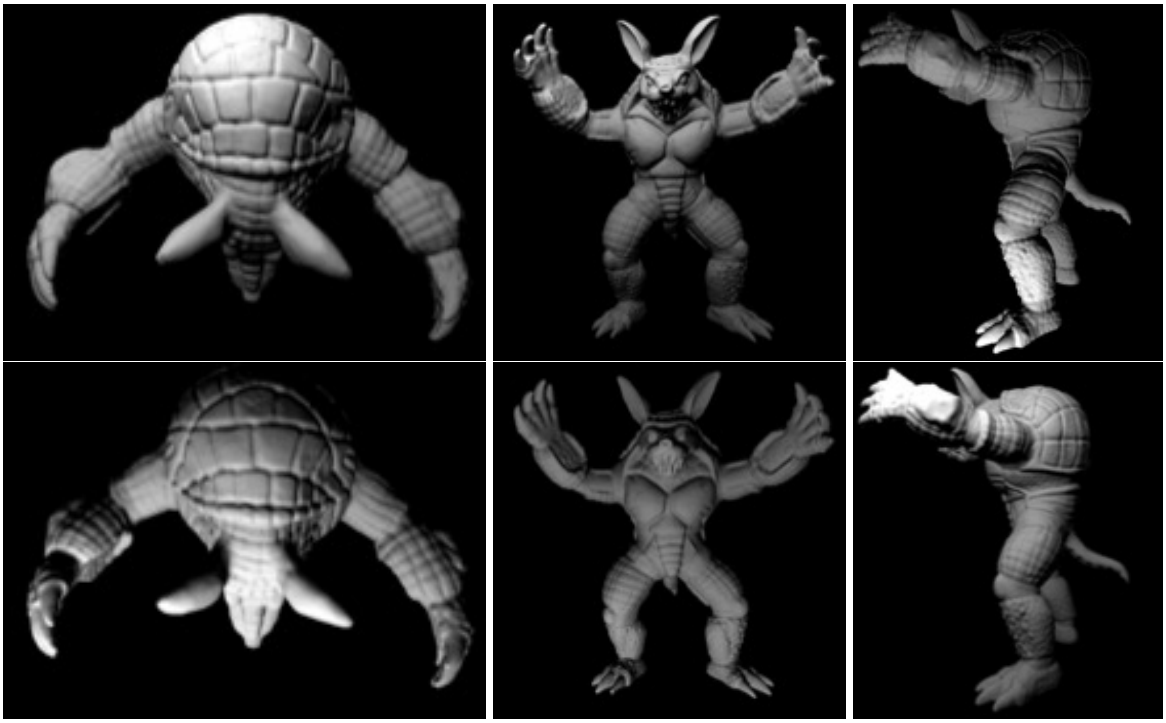
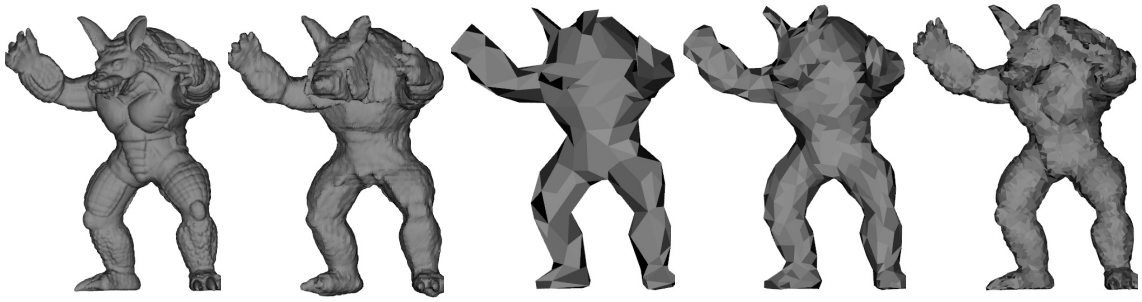


Figure 6.13 Synthetic data samples featuring near lighting, perspective deformation and self-occlusions.

The proposed algorithm is evaluated on synthetic and real data. Comparison is against Park et al. [144] using the code from their website. It is worth to mentioning that in contrast to the proposed method, their state-of-the-art approach for MVPS is based on a fully un-calibrated PS model.

6.7.1 Synthetic test cases

Synthetic experiments were performed using the "Armadillo" model. This object has non-trivial geometry that generates several self-occlusions as well as cast shadows. The virtual object was scaled to have approximate radius 20mm and the virtual camera with a focal length of 6mm, was placed in several positions on a sphere of radius 45mm around the



(a) Ground truth (b) Visual Hull (c) Initial estimates: 500,1500,10K triangles with 0% noise.

Figure 6.14 Synthetic data experiment- initial geometry estimates used for initialising the MVPS optimisation. The initial estimates of variable quality in (c) were generated by subsampling the ground truth (a), using the edge collapse decimation function of Meshlab.

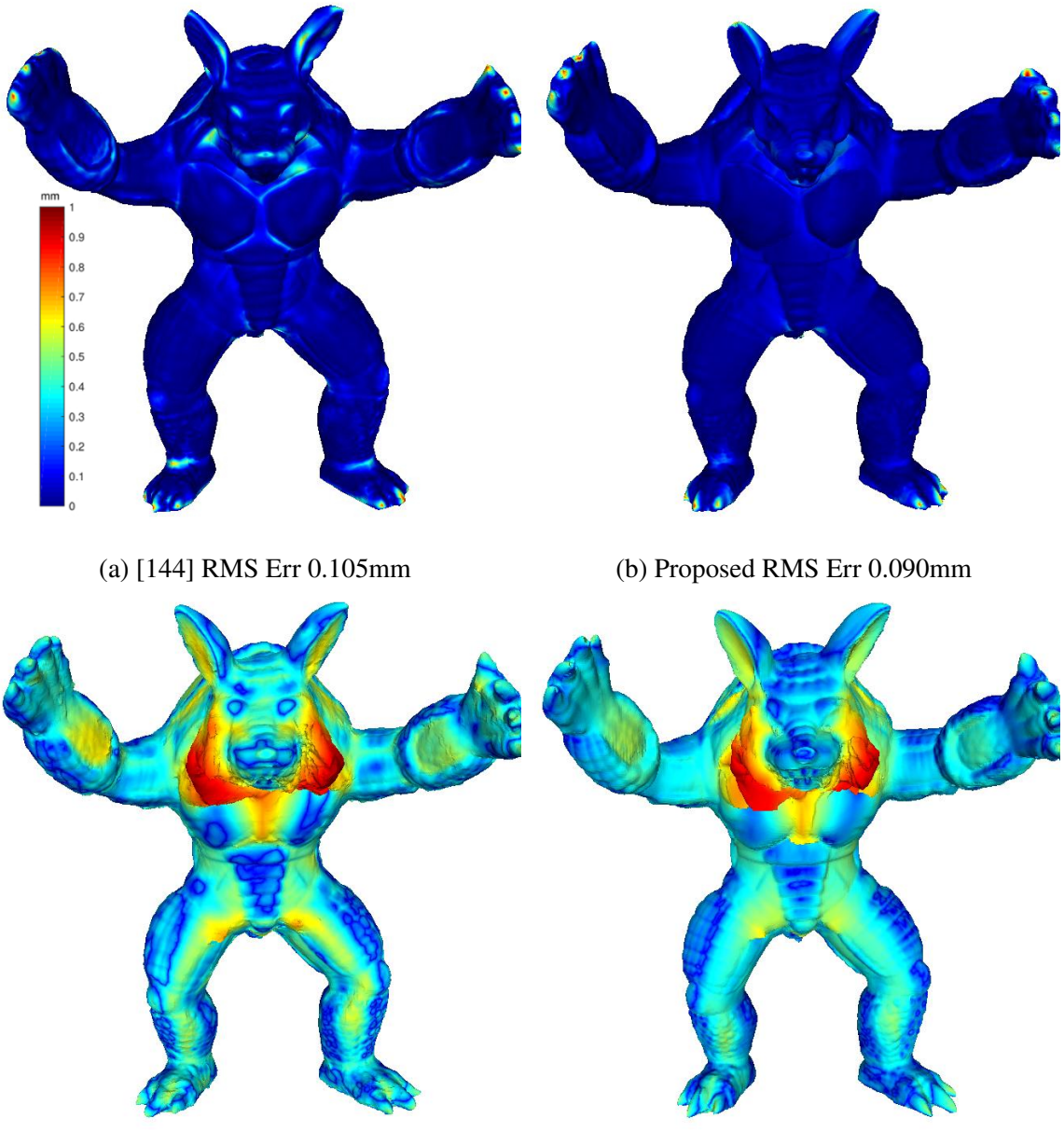
object. Thereafter, 12 views with 8 images each of resolution 1200x800x24bits per pixel (see Figure 6.13) were rendered.

For the initial geometry estimate the following two different strategies were employed: The first one, used a sub-sampled version of the ground truth mesh with some added noise. This was done in order to minimise the influence of systematic errors in the initial estimate on the final result and to evaluate the MVPS algorithm alone. The original 150k triangle mesh was subsampled using the quadric edge collapse decimation function of Meshlab. Five different meshes were created with the number of triangles ranging from 250 to 30K. Some of these meshes are shown in Figure 6.14. In addition, for each of these meshes Gaussian noise was added to the vertex coordinates with a standard deviation of 0%, 5%, 10% of the average triangle size. This in turn creates a total of 15 different experiments.

The second strategy adopted for generating an initial estimate of the geometry was to use the object's visual hull. This was calculated with naive voxel carving on a 128x128x128 voxel grid. Note that, as the previous set of synthetic experiments show, (see Table 6.1), there is no need for an initial estimate of very high resolution, hence the low resolution grid for computing the visual hull.

The evaluation metric is the RMS Hausdorff distance to the ground truth (computed with Meshlab). This is computed by calculating the distance of each vertex of the mesh in question, with the closest triangle of the reference mesh. The complete results for all the experiments are shown in Table 6.1. Figures 6.15, 6.16 and 6.17 also show a few sample renderings of the results.

The proposed approach outperforms [144] in all experiments, although in the high-triangles initial estimates and, low-noise cases the error is imperceptibly small and the difference between the proposed approach and [144] is probably statistically insignificant.



(a) [144] RMS Err 0.105mm

(b) Proposed RMS Err 0.090mm

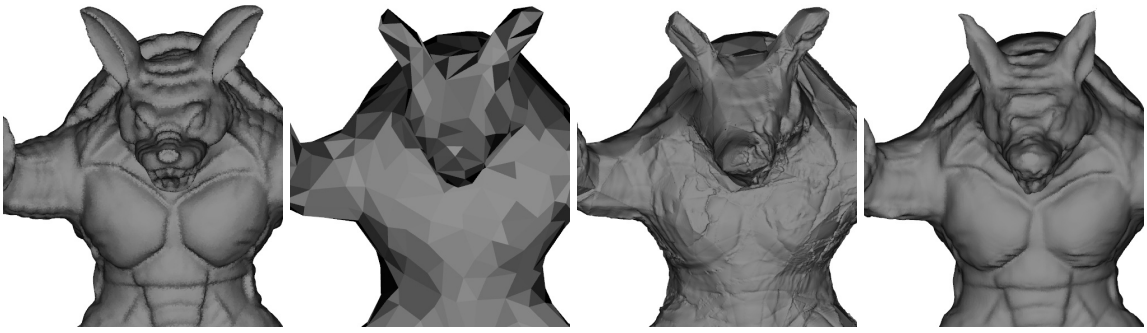
(c) [144] RMS Err 0.445mm

(d) Proposed RMS Err 0.386mm

Figure 6.15 Evaluation using the 1500-triangles mesh initial estimate ((a) and (b)) and the visual hull initial estimate ((c) and (d)). The colour coding shows error (Hausdorff distance computed with Meshlab) with respect to the ground truth in mm.

6.7.2 Real data

The real datasets include a marble Buddha statue, a plaster bust of Queen Elizabeth (see Figure 6.1), and a combined scene with a swede next to a porcelain cup containing a small tree branch (see Figure 6.1). The images were acquired while moving the objects on a



(a) Ground truth (b) Initial estimate (c) [144] (d) Proposed

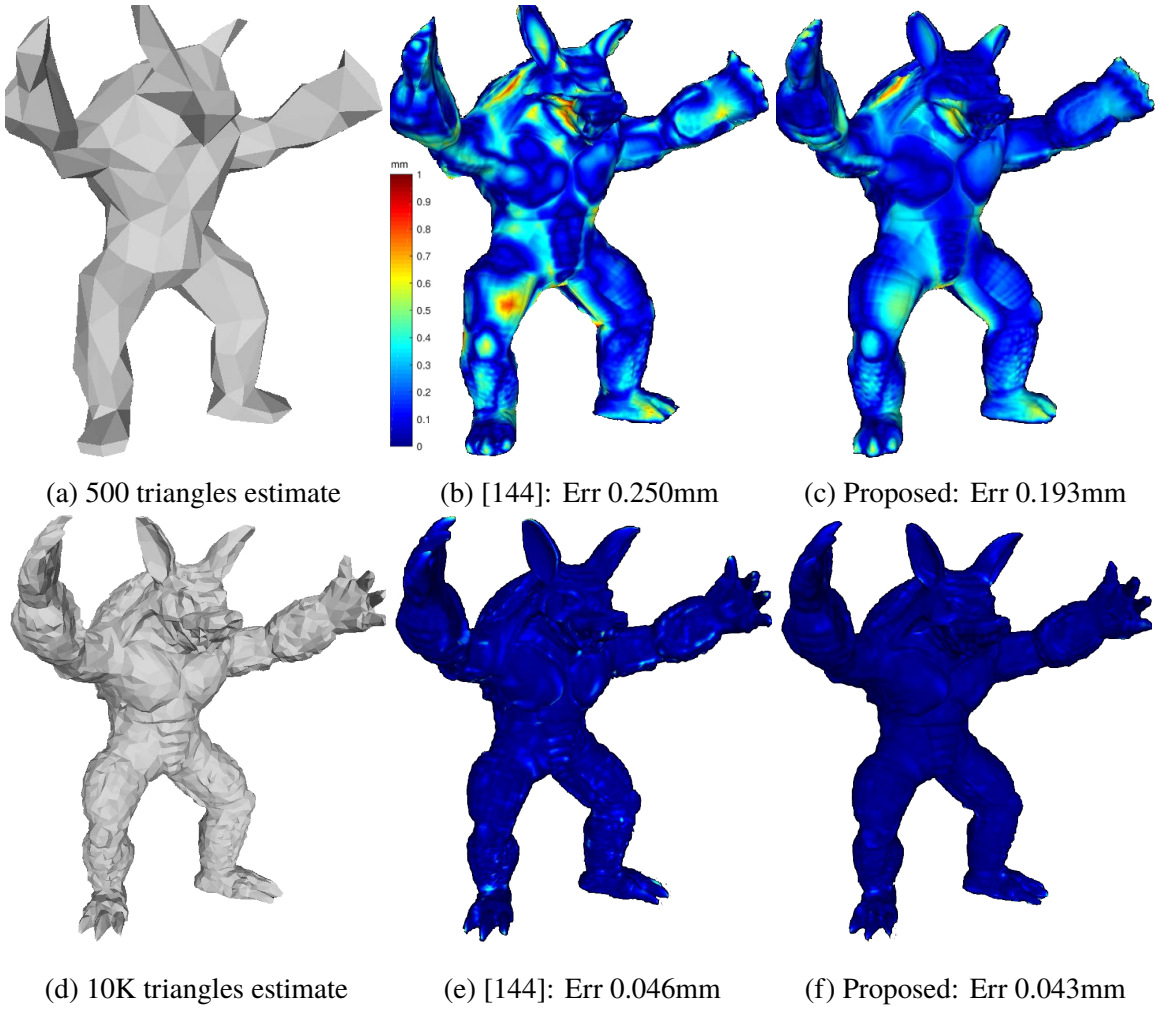
Figure 6.16 Close-up rendering of: the ground truth (a), the 1500 triangles initial estimate with 0% noise case (b) (column 3 of Table 6.1), as well as the reconstructions obtained with [144], (c) and the proposed method (d).

Table 6.1 A quantitative evaluation based on the quality of the initial estimate . Errors are in mm. Noise added to the vertex positions, and the magnitude is relative to the average triangle size.

Experiment		Triangle Number					Visual Hull
Method	Noise	250	500	1500	10K	30K	69K
[144]	0%	0.245	0.141	0.105	0.029	0.025	0.445
	5%	0.290	0.172	0.119	0.036	0.029	-
	10%	0.393	0.250	0.153	0.046	0.031	-
Proposed	0%	0.203	0.114	0.090	0.026	0.023	0.386
	5%	0.234	0.137	0.104	0.033	0.024	-
	10%	0.321	0.193	0.131	0.043	0.028	-

turntable, essentially generating a few circular trajectories around them (see Figure 6.2). However, neither the SFM-MVS pipeline, nor the MVPS algorithms used the circular motion information at all.

The multi-view data have been processed using VisualSFM [193, 194] and PMVS [49] for getting camera rotation and translation between the PS views as well as a low quality reconstruction to use as initial estimate. In addition, a few more images were captured in between the PS sequences (with neutral illumination) in order to make SFM more robust with respect to a too small overlap between images. Moreover, some noisy regions and background points far away from the scenes of interest were manually removed. Finally, Poisson reconstruction [84] was performed with a low level setting so as the initial estimate contains continues surfaces (and not point clouds) and thus be able to generate an initial



(a) 500 triangles estimate (b) [144]: Err 0.250mm (c) Proposed: Err 0.193mm

(d) 10K triangles estimate (e) [144]: Err 0.046mm (f) Proposed: Err 0.043mm

Figure 6.17 Additional renderings from Table 6.1. 10% noise is included in the initial estimates shown in the first column. The colour coding shows error with respect to the ground truth in mm.

SDF estimate (using the distance transform function of Matlab). As Table 6.1 suggests, the proposed method does not need a very accurate initial estimate.

The proposed approach outperforms [144] in all three datasets (Figures 6.19 and 6.20) and is able to recover more detailed surfaces. One of the main reasons for the inability of [144] to recover very detailed reconstructions is because it is truly limited by the very low quality of the initial estimates: they parameterise the initial surface into a 2D domain and then perform a single refinement step. In fact, triangulation artefacts can be seen at their reconstruction of the forehead of the Queen (Figure 6.19 top right). The reason for their performing a single step reconstruction is the fact that their 2D parameterisation and visibility estimation get very computationally expensive as a function of the resolution of the

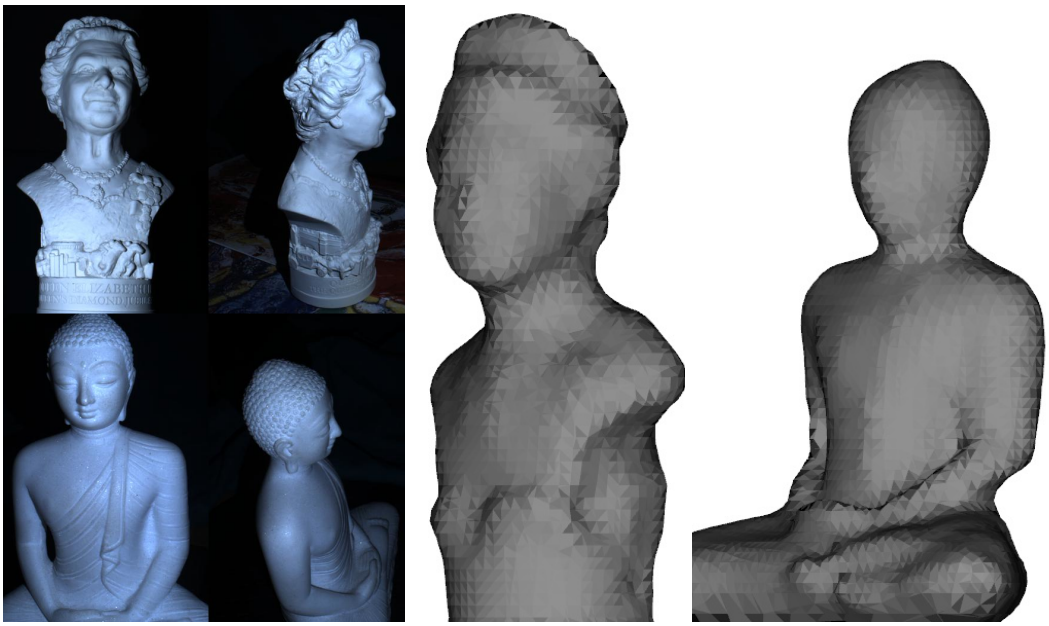


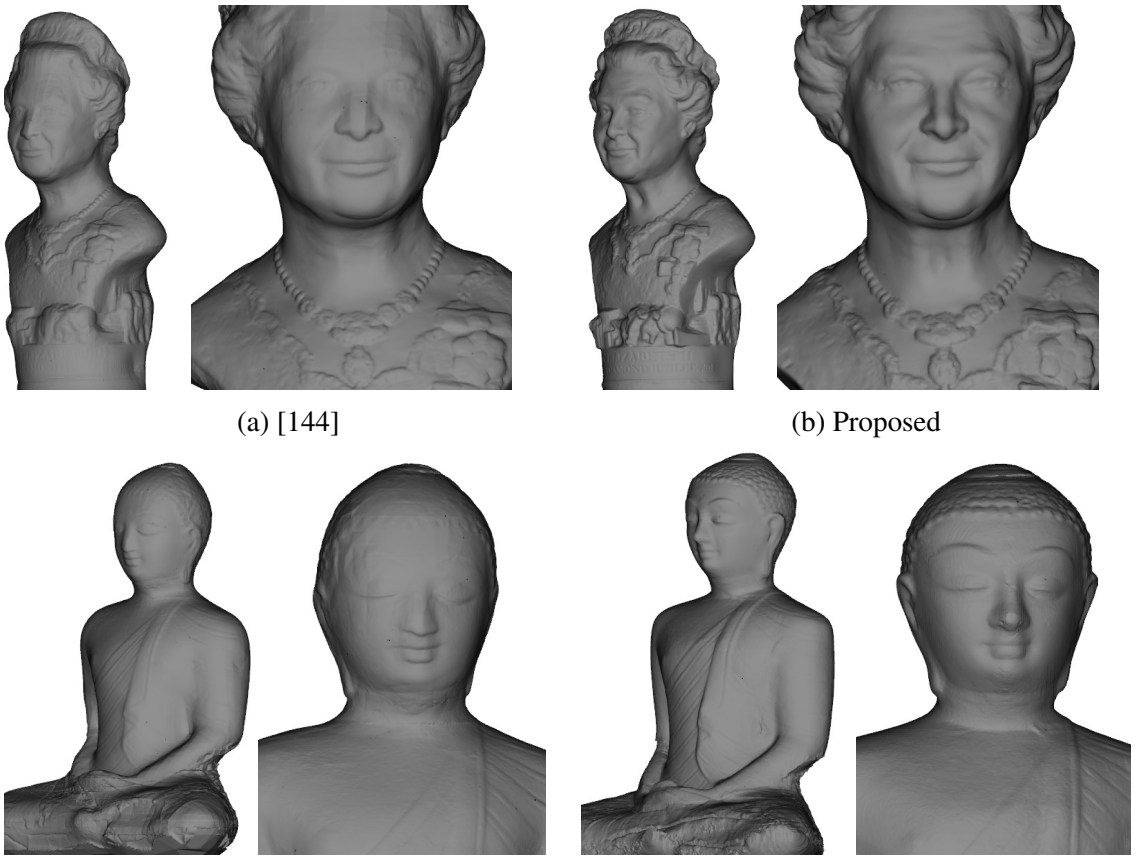
Figure 6.18 Real data: 2/96 photometric stereo images (12 views with 8 lights in each view were used) and initial geometry estimate obtained with MVS. This initial estimates are only 8k and 11k triangles for the Queen and Buddha datasets respectively.

mesh. In contrast, the proposed volumetric approach naturally handles multiscale estimation by arranging the voxels into an octree structure and so it is easy to repeatedly refine the volume estimate until the voxels have a 1-1 correspondence with image pixels. Finally, we note that the datasets presented in [144] have much higher quality initial estimates (their "AccordionMan" starts from 140k triangles versus the 8k triangles initial estimate for the "Queen" dataset).

6.8 Discussion

A volumetric parameterisation based on the signed distance function for the MVPS problem was presented. Very high precision is achieved by using an octree implementation for processing and ray-tracing the volume on a tree. While considering Photometric Stereo images, the proposed fully differential formulation is albedo independent as it uses the irradiance equation ratio approach for the near-field Photometric Stereo presented in Section 3.3.4.

The main limitation of the proposed approach is the inability to cope with missing big portions of scene (this also true for most competing approaches e.g. [144, 209, 198]). For example, if the initial reconstruction is missing the hands of the Armadillo, they will not be recovered. This can potentially lead to to degradation of quality for the rest of the body as



(a) [144] (b) Proposed

(c) [144] (d) Proposed

Figure 6.19 Qualitative evaluation on real data set of Figure 6.18. The proposed approach outperforms [144] and generates more detailed reconstructions.

well, as the missing parts will lead to sub-optimal estimates of visibility. The theoretical justification for this is that the core assumption $\boldsymbol{N}(\boldsymbol{X}) = \nabla d(\boldsymbol{X})$ is only exactly true for points on the true surface, i.e. $d(\boldsymbol{X}) = 0$. Assuming continuity, it is assumed that $\boldsymbol{N}(\boldsymbol{X}) \approx \nabla d(\boldsymbol{X})$ if $d(\boldsymbol{X}) \approx 0$ so the differential approach is performed under the assumption that the set of voxels is relatively close to the true surface.

The main drawback of the proposed method compared to mesh parameterisation techniques (e.g. Park et al. [144]) is the elevated memory requirements. Even though the octree implementation minimises the number of voxels required, it is inevitable to need a few voxels per each potential surface point[3] . In addition, the use of the variational optimisation is also memory expensive as the matrix encoding the neighbouring information about voxels needs to be stored in memory as well.

[3] As the surface is the zero crossings of the SDF, at least one pair of opposite signed values are required per surface point.

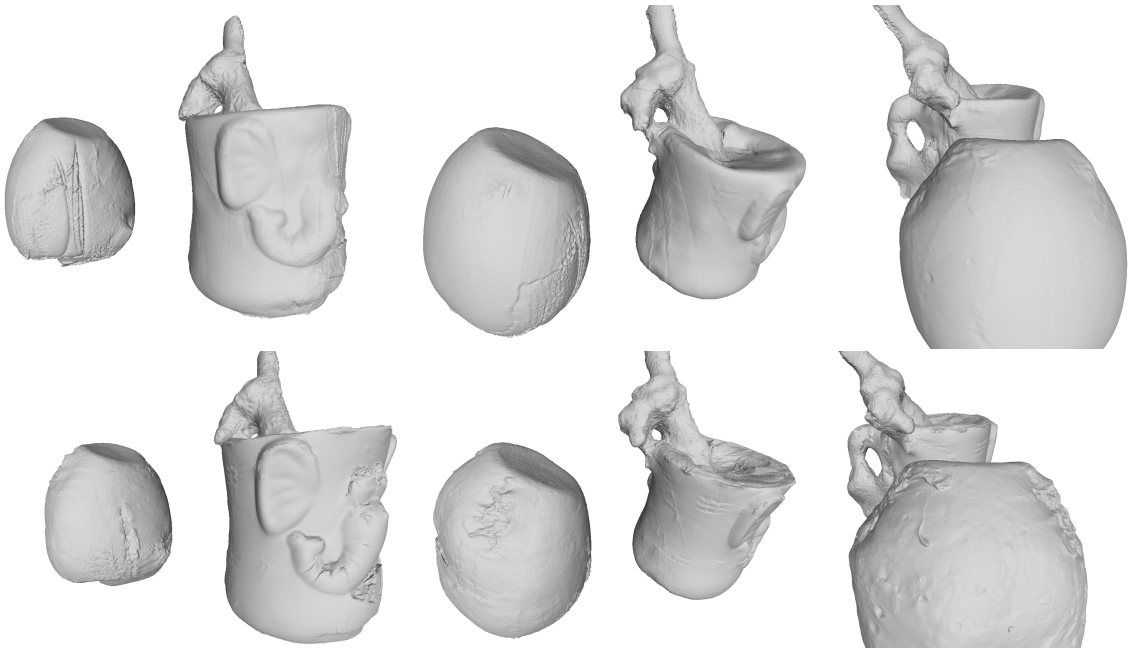


Figure 6.20 Qualitative evaluation using the multi-object scene of Figure 6.1 . Line 1 [144], line 2 proposed. The proposed approach is superior to [144] in simultaneously recovering the rough surface of the swede as well as the smooth surface of the cup.

6.8.1 Future Work

There are several directions in which the work presented in this chapter can be extended:

More advanced PS modelling The proposed MVPS parameterisation is quite minimal compared to the rest of the thesis in terms of the PS modelling features. Indeed, any kind of specular reflection is ignored, the ambient light is considered negligible and the light sources are assumed fully calibrated. Nonetheless, extending the proposed model to include all these features from Chapters 4 and 5 is pretty straightforward and most of the difficulty lies in its implementation, as the computational requirements will be highly elevated.

Independence from SFM As the proposed method is crucially reliant on SFM algorithms for calibrating camera positions and some MVS stereo for an initial geometry estimate, it is susceptible to failure when SFM fails, because of, for example, too few point features or too big a camera movement between views, which make standard matching based on for example SIFT (Lowe [109]) to fail. The naive way to eliminate the dependence on SFM would be to perform 2D PS from different vantage points and then attempt to recover camera motion using iterative closest point (ICP) [16]. However, as discussed in Section 6.3.1, PS reconstructions from different viewpoints are usually inconsistent, and thus ICP is not

expected to be very reliable. A more promising future direction would be to try to perform matching using frequency domain features or a feature like the Laplace-Beltrami operator (divergence of the gradient). This is directly computable from the PS image data in each view (without the need for numerical integration /variational solver) and could possibly be used to track camera motion.

Another potential way of eliminating the dependence on MVS would be to use the rank constraint Equation 6.10 in order to obtain a geometry estimate. Indeed, this constraint encodes the compatibility of the SDF level-set for different views, with a low condition number of B (Λ_1/Λ_3) signifying a point far away from the surface. Thus, this could be used as a volume cost function and incorporated into volumetric graph-cut optimisation like the one of Vogiatzis et al. [184].

Chapter 7

Applications and future work

7.1 Findings

The findings presented in this dissertation have contributed towards the attainment of a more refined, practical and viable 3D technology. The resultant outcomes of this work can therefore be summarised as follows:

Ambient light independence Chapter 4 demonstrated how the Photometric Stereo model can be extended in order to deal with arbitrary ambient lighting while retaining all the realistic modelling assumptions of the state-of-the-art approaches namely perspective viewing geometry, nearby point light sources and specular reflection. This parameterisation allows acquisition in an open environment and eliminates the need of a dark room for acquisition.

Semi-calibrated formulation The need for a priori knowledge of the light source brightness and attenuation characteristics was relaxed, as illustrated in Chapter 5. It was shown that these parameters can be estimated through an alternating optimisation procedure along with the geometry and the specularity coefficients. This extension allows for quick acquisition with inexpensive LEDs that exhibit unpredictable illumination characteristics like flickering.

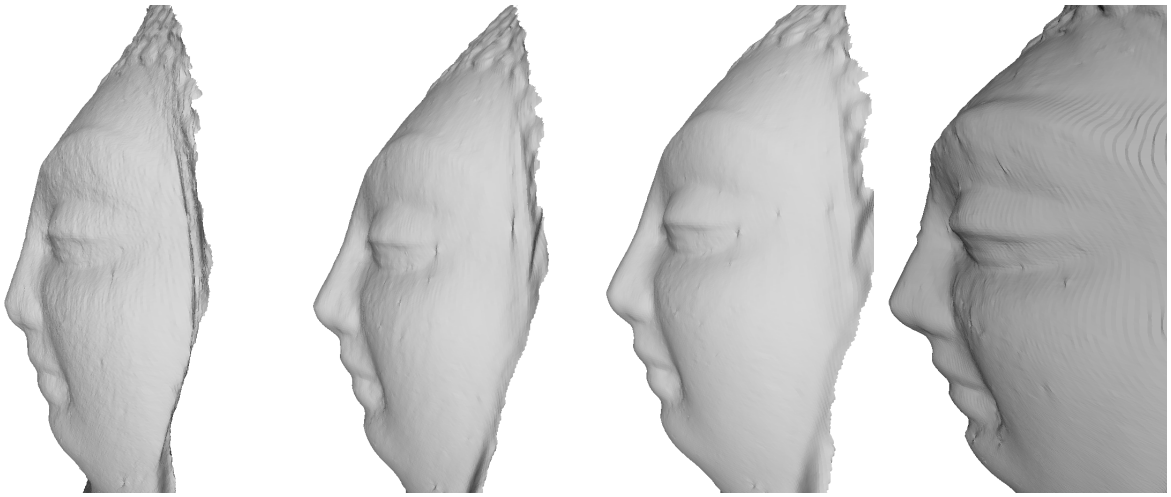
Multi-view parameterisation A novel volumetric differential parameterisation based on the signed distance field was proposed in Chapter 6. This formulation allows one to tackle the multi-view Photometric Stereo problem in a simple unified differential model where information from multiple viewpoints and illumination is fused in a single optimisation. Hence, scenes with multiple discrete objects can be reconstructed.

7.2 Applications

The theoretical framework developed in this thesis is applicable in generalized theoretical and practical settings. One such practical and another such theoretical application is presented below.

7.2.1 GPU implementation

A practical application for the framework of this thesis is the implementation of a PS scanning system that allows for very fast acquisition and reconstruction. As a number of computations can be performed independently for each pixel, the reconstruction algorithms presented can be implemented on a GPU, assigning different computation threads to different pixels. These parallelisable operations include calculation of a and $\boldsymbol{W}$ fields for the lighting model, $\boldsymbol{b}$ and s fields for the PDE as well as the computation of the specularity parameter c.



(a) 8 images, 3.7s (b) 12 images, 3.8s (c) 24 images, 9.3s (d) 8 images, 155s

Figure 7.1 Comparison of the simplified GPU implementation with the full model for the ambient light problem and the Buddha dataset of Figure 4.1. (a) to (c) GPU reconstructions considering all pairs and (d) is the CPU Order 2 method considering all 6-tuples of images.

The implemented version is mostly based on the semi-calibrated framework as outlined in Chapter 5. In addition, the ambient light model presented in Chapter 4 is also implemented. Finally, the system also includes a mean depth/scale estimation capability based on the final reconstruction reprojection error as discussed in Section 6.6. The most important changes between the GPU version and the rest of the thesis was the usage of the ℓ_2 version of the variational solver (Section 3.7.2) instead of the ℓ_1. The reasoning for this simplification is the highly reduced memory requirements as explained in Section 3.7.2. In addition, the shadow

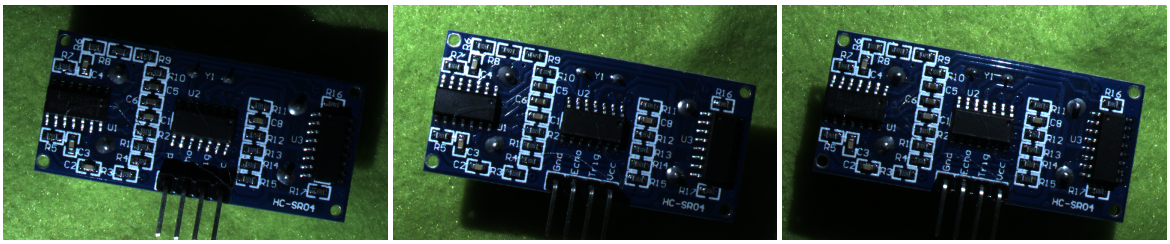
map and attenuation map steps of Chapter 5 are also avoided as these are very computationally expensive with not much increase in the reconstruction quality.

One major advantage of the GPU implementation compared to its CPU counterparts is the ease with which datasets with large number of images is dealt with. In fact, one major disadvantage of the ratio method is that the computational complexity of computing the PS model (PDE coefficients) is scaling as $O(n^2)$ for a number of images n. This is because pairs of images need to be considered. The computational complexity is even worse for the ratio of differences method presented in Chapter 4, which has $O(n^4)$ complexity for the Order 1 method and $O(n^6)$ for that of Order 2. This problem is alleviated through GPU implementation. Even though the complexity is the same, the GPU allows iterating through all n-tuples of images very quickly even when this number of images is large, as this is a completely parallelisable operation. This is experimentally verified in Figure 7.1.

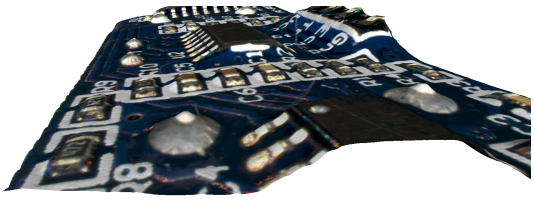
As it can be seen in Figure 7.1, even though the Order 2 full model better preserves the depth of the head and gets a better overall reconstruction, the computation time is significantly higher. In fact, the more interesting comparison is between Figure 7.1a and Figure 7.1c: increasing the number of images from 8 to 24 increases the number of quadruples from 70 to 10626 but the total computation time only increases from 3.7 to 9.3 seconds. That is because the $\boldsymbol{b}$ and s field calculation is fully parallelisable taking around 0.04ms per quadruple per iteration[1]. The more computationally expensive part (dominating the 3.7s seconds of the 8 images case) is the system solution and this is independent of the number of images for ℓ_2 solver. In addition, it is noted that performing an ℓ_1 computation on the 24 images dataset with all possible quadruples would be computationally intractable as it would be required to store all 10626 residuals per pixels in memory.

Finally, the ability of this implementation to handle datasets with a large number of images, makes it very suitable for the task of scanning specular objects, a specific example being that of circuit inspection. As discussed in Section 2.1, the distinctive characteristic of specular objects is a good portion of the reflected light comes in a single direction. As a result, a lot of light sources are required in order to make sure that each pixel has a well constraint set of equations since most image data is completely black. Figure 7.2 shows the results on a circuit dataset. These reconstructions were generated with a 30 images dataset using the GPU implementation of the semi-calibrated model; the ℓ_1 version of the variational solver would run out of memory trying to store all residuals for the approximately 900K pixels in the image domain in all $\binom{30}{2} = 435$ pairs; in contrast the GPU version of the ℓ_2 solver produces the reconstructions of Figure 7.2 in 19.4 seconds.

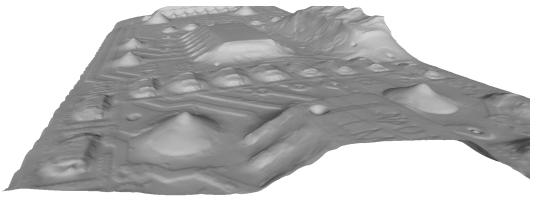
[1]These timings are reported for an NVIDIA GTX 780 on the Buddha dataset with image resolution 800x600 and around 180K pixels in the reconstruction domain.



(a) 3/30 sample images.



(b) Circuit textured.



(c) Circuit geometry.

Figure 7.2 3D reconstruction results on the Printed Circuit Board (PCB) inspection task. This is a fairly challenging task due to the specular reflection of the metallic parts. As a result, 30 images were used and the optimisation needed to be initialised with $c = 0.2$ for the specularity parameter (see Section 3.4.2) instead of the usual $c = 1$ (Lambertian). The reconstruction recovers a large amount of details in both geometry and albedo and does not suffer from any significant low frequency deformation. The only visible artefact is the smoothing at depth discontinuities at the point of contact of the integrated circuits, which is expected due to the use of the ℓ_2 solver.

7.2.2 Shape from polarisation

A theoretical application of the proposed formulations of the thesis is in the Shape from Polarisation (SFP) problem. This was done in collaboration with Roberto Mecca in the work Mecca et al. [116]. SFP was introduced in Koshikawa [91] and has a very large literature, with important results being those of Atkinson and Hancock [7] and Smith et al. [170]. Only a brief explanation is presented here.

SFP involves taking multiple images from the same point of view with constant illumination but with a polarisation filter in front of the camera. The polarisation filter is rotated and this rotation causes a sinusoidal variation in the image intensity. Crucially, this variation is a function of the surface normal as well as material properties.

Mecca et al. [116] and the extension at Logothetis et al. [106] present a differential formulation for the SFP problem which has the form

$$\boldsymbol{b}_{pol} \cdot \nabla z = 0, \tag{7.1}$$

where $\boldsymbol{b}_{pol}$ is independent of illumination and can easily be computed from the polarisation data. Equation 7.1 trivially reveals the level-set of the surface. This is true for diffuse and specular objects and even works for semi-transparent objects like glass. This was experimentally verified as seen in Figure 7.3.

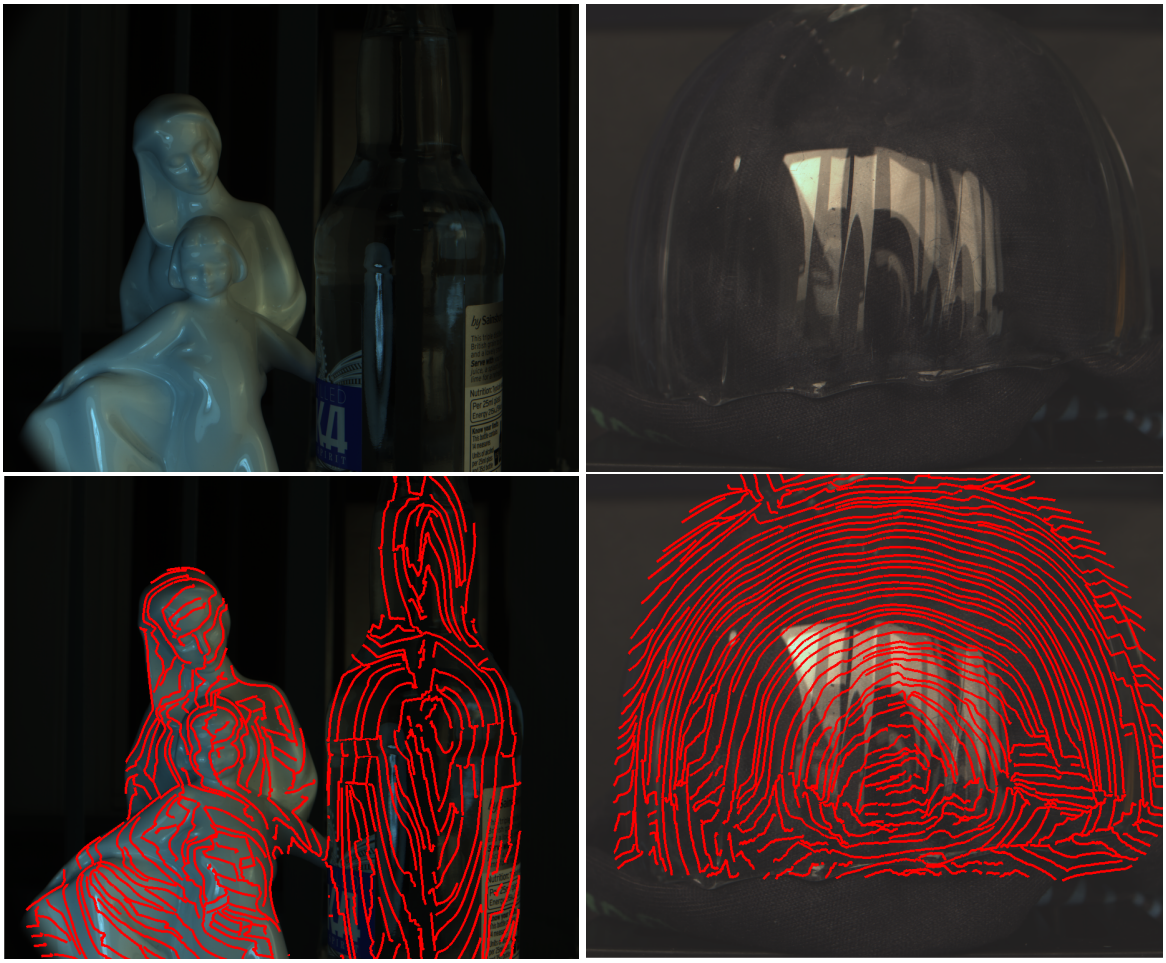


Figure 7.3 On the top: two sample images of objects obtained using a polarisation filter. On the bottom: the respective level-sets of these objects obtained using the method of Logothetis et al. [106].

The most important property of Equation 7.1 in the context of this thesis is that it has the same form of the unifying PS Equation 3.18 $\boldsymbol{b}_{ps} \cdot \nabla z = s$ and hence allows stacking these 2 equations to subsequently obtain a unified variational problem of the form (including the standard Tikhonov regulariser term):

$$\min_{z} \left\| \begin{bmatrix} \boldsymbol{b}_{ps} \\ \boldsymbol{b}_{pol} \end{bmatrix} \cdot \nabla z - \begin{bmatrix} s_{ps} \\ 0 \end{bmatrix} \right\|_{\ell_2} + \lambda \left\| z - z_0 \right\|_{\ell_2} . \tag{7.2}$$

As a technical note, ℓ_2 was chosen in Equation 7.2 as the aim of [116] was to present a minimal solution; thus ℓ_1, which aims at dealing with sparse outliers is not really applicable in the minimal solution case. In fact, the merging of polarisation data allows to get a well-posed problem with using a single pair of calibrated photometric stereo images. Equation 7.2 is then discritised and solve with the method described in Section 3.7. Some real data results are shown in Figure 7.4.

(a) Photometric stereo pairs of images used.

(b) Minimum data reconstructions.

Figure 7.4 Shape from Polarisation data and reconstructions. The polarisation images are not shown as for most pixels the degree of polarisation is only a small percentage of the total value making this difference imperceivable. The reconstructions shown at (b) are using a minimum of possible data and thus are of inferior quality to those that one would get using multiple PS images.

7.3 Limitations and future work

Important limitations that are true for all the work presented in this thesis are the following:

Strict reliability on the assumed reflectance model Although the combination of Blinn-Phong and Lambertian BRDF introduced in Section 3.4.2 has been shown to be able to deal with heterogeneous objects, even with metallic like the circuit in Figure 7.2, it is not reliable in all possible reflectance cases, because the Blinn-Phong model is not physically based.

Dependence on continuity and differentiability A very important assumption throughout the thesis has been that surfaces can be parameterised as differentiable depth maps or differentiable signed distance fields (in Chapter 6). Of course, in reality this is not always true with notable exceptions being sharp edges and depth discontinuities at occlusion boundaries. The strategy throughout the thesis has been to treat these imperfections as sparse outliers and try to maximise their robustness. In addition, all of the data presented (even the multi-object scene in Figure 6.1) have been preprocessed with a background subtraction step[2]. Hence, applying the developed method to the task of reconstructing all visible objects in a scene is not trivially possible.

Non-provable convergent optimisations All of the PS models presented correspond to non-convex optimisation problem. Therefore, none of the algorithms presented are guaranteed to converge to the global optimum, and even when they converge, there is no guarantee that the rate of convergence will be reasonable. However, the extensive experimental evaluation has shown that this issue is not applicable in practice and these algorithms would converge to reasonable reconstructions with no need for special initialisation or priors. A notable exception is the circuit dataset in Figure 7.2; as this is very specular, the optimisation needed to know that a priori. In addition, the multi-view framework in Chapter 6 needs to be initialised from an MVS estimate which is relatively close to the real surface and would not converge from a very generic initialisation like a sphere/cube.

The need for some control of lighting Finally, it has to be emphasised that any PS method requires controlled variation in illumination, with at least 2 images per point of view for a well constrained problem (avoiding all the SFS ambiguities discussed in Section 2.3). This in turn makes PS inapplicable for use with consumer cameras and mobile phones.

7.3.1 Future work

There are several ways in which the overall directions of the work can be further extended. These include:

Additional 3D reconstruction cues An interesting theoretical but also practically useful future research direction is the inclusion of additional cues. For example, a very promising method is the Shape From Defocus (Favaro et al. [45]) which can extract 3D geometrical features by changing the focus distance of the lens. In general, the inclusion of extra cues can

[2]using the semi-manual graph-cut, OpenCV implementation https://docs.opencv.org/3.1.0/d8/d83/tutorial_py_grabcut.html

widen the field of applicability to many more general surfaces, effectively dealing with the first limitation discussed above.

Multi-view Polarisation To add to the previous point, it is noted that the polarisation Equation 7.1 could be incorporated into the multi-view framework of Chapter 6. This is a natural extension to using just two lights fusion presented in Mecca et al. [116] could ultimately lead to full 3D reconstruction of even semi-transparent objects like a glass bowl of Figure 7.3 (as level-sets from two different viewpoints can be used to find a surface).

Reflectance Capture Almost all of the work in the thesis has focused on recovering the 3D structure of objects. However, various applications such as heritage preservation, 3D graphics and virtual reality require accurate capture of the reflectance properties of the objects, in order to achieve faithful re-renderings of these items. Some preliminary work has been the simultaneous recovery of albedo and specularity coefficients in Chapter 5, however, a fully realistic rendering of these objects would require further work.

7.4 Final remarks

The Photometric Stereo method has always been open to interesting theoretical research due to the complicated light reflection models developed, as well as the differential relationship between surfaces and normals, which allows posing the problem in a PDE form. Having specifically aimed at making a practical reconstruction system, this dissertation has advanced the theoretical modelling of Photometric Stereo. By combining manipulations of PDEs and variational methods with detailed engineering choices like oct-trees and GPU implementation, along with the use of a custom made hardware setup, the 3D reconstructions attained verify that a small step in the right direction has been taken.

Bibliography

[1] Abrams, A., Hawley, C., and Pless, R. (2012). Heliometric stereo: Shape from sun position. In *12th European Conference on Computer Vision (ECCV)*, pages 357–370.

[2] Ackermann, J., Langguth, F., Fuhrmann, S., and Goesele, M. (2012). Photometric stereo for outdoor webcams. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 262–269.

[3] Ackermann, J., Langguth, F., Fuhrmann, S., Kuijper, A., and Goesele, M. (2014). Multi-view photometric stereo by example. In *2014 2nd International Conference on 3D Vision (3DV)*, volume 1, pages 259–266.

[4] Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., and Szeliski, R. (2011). Building rome in a day. *Communications of the ACM*, 54(10):105–112.

[5] Alldrin, N. G. and Kriegman, D. J. (2007). Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *2007 IEEE International Conference on Computer Vision (ICCV)*, pages 1–8.

[6] Anderson, R., Stenger, B., and Cipolla, R. (2011). Color photometric stereo for multicolored surfaces. In *2011 IEEE International Conference on Computer Vision (ICCV)*, pages 2182–2189.

[7] Atkinson, G. A. and Hancock, E. R. (2007). Shape estimation using polarization and shading from two views. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(11).

[8] Bähr, M., Breuß, M., Quéau, Y., Boroujerdi, A. S., and Durou, J.-D. (2017). Fast and accurate surface normal integration on non-rectangular domains. *Computational Visual Media*, 3(2):107–129.

[9] Barron, J. T. and Malik, J. (2015). Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(8):1670–1687.

[10] Barsky, S. and Petrou, M. (2003). The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(10):1239–1252.

[11] Basri, R. and Jacobs, D. (2001). Lambertian reflectance and linear subspaces. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 383–390.

[12] Basri, R., Jacobs, D., and Kemelmacher, I. (2007). Photometric Stereo with General, Unknown Lighting. *International Journal of Computer Vision (IJCV)*, 72(3):239–257.

[13] Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.

[14] Belhumeur, P. N., Kriegman, D. J., and Yuille, A. L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision (IJCV)*, 35(1):33–44.

[15] Beljan, M., Ackermann, J., and Goesele, M. (2012). Consensus multi-view photometric stereo. In *Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*, pages 287–296.

[16] Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-D shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pages 586–607.

[17] Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. In *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 192–198.

[18] Bouguet, J.-Y. (2015). Camera calibration toolbox for matlab. Retrieved on 1/4/2016 from http://www.vision.caltech.edu/bouguetj/calib_doc/.

[19] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, New York, NY, USA.

[20] Breuß, M., Cristiani, E., Durou, J.-D., Falcone, M., and Vogel, O. (2012). Perspective Shape from Shading: Ambiguity Analysis and Numerical Approximations. *SIAM Journal on Imaging Sciences*, 5(1):311–342.

[21] Brostow, G. J., Hernández, C., Vogiatzis, G., Stenger, B., and Cipolla, R. (2011). Video normals from colored lights. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(10):2104–2114.

[22] Bruckstein, A. M. (1988). On shape from shading. *Computer Vision, Graphic, and Image Processing (CVGI)*, 44(2):139–154.

[23] Candès, E. J., Wakin, M. B., and Boyd, S. P. (2008). Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905.

[24] Chandraker, M., Bai, J., and Ramamoorthi, R. (2011). A theory of differential photometric stereo for unknown BRDFs. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2505–2512.

[25] Chandraker, M., Bai, J., and Ramamoorthi, R. (2013). On Differential Photometric Reconstruction for Unknown, Isotropic BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(12):2941–2955.

[26] Chandraker, M. K., Agarwal, S., and Kriegman, D. J. (2007). Shadowcuts: Photometric stereo with shadows. In *2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8.

[27] Chen, G., Han, K., and Wong, K.-Y. K. (2018). Ps-fcn: A flexible learning framework for photometric stereo. In *15th European Conference on Computer Vision (ECCV)*.

[28] Cho, D., Matsushita, Y., Tai, Y.-W., and Kweon, I. (2016). Photometric stereo under non-uniform light intensities and exposures. In *14th European Conference on Computer Vision (ECCV)*, pages 170–186.

[29] Chung, H. J. and Jia, J. (2008). Efficient photometric stereo on glossy surfaces with wide specular lobes. In *2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8.

[30] Clark, J. J. (1992). Active photometric stereo. In *1992 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 29–34.

[31] Clark, J. J. (2010). Photometric stereo using LCD displays. *Image and Vision Computing*, 28(4):704–714.

[32] Cook, R. L. and Torrance, K. E. (1982). A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*, 1(1):7–24.

[33] Courteille, F., Crouzil, A., Durou, J. D., and Gurdjos, P. (2004). Towards shape from shading under realistic photographic conditions. *17th International Conference on Parttern Recognition (ICPR)*, pages 277–280.

[34] Curless, B. and Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312.

[35] Daniel, P. and Durou, J. D. (2000). From deterministic to stochastic methods for shape from shading. *Asian Conference on Computer Vision (ACCV)*, pages 187–192.

[36] Davis, P. A. and Soderblom, L. A. (1984). Modeling crater topography and albedo from monoscopic viking orbiter images: 1. methodology. *Journal of Geophysical Research: Solid Earth*, 89(B11):9449–9457.

[37] Debevec, P. E., Hawkins, T., Tchou, C., Duiker, H., Sarokin, W., and Sagar, M. (2000). Acquiring the reflectance field of a human face. In *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 145–156.

[38] Debevec, P. E., Wenger, A., Tchou, C., Gardner, A., Waese, J., and Hawkins, T. (2002). A lighting reproduction approach to live-action compositing. *ACM Transactions on Graphics (TOG)*, 21(3):547–556.

[39] Delaunoy, A. and Prados, E. (2011). Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3D reconstruction problems dealing with visibility. *International Journal of Computer Vision (IJCV)*, 95(2):100–123.

[40] Durou, J., Falcone, M., and Sagona, M. (2008). Numerical methods for shape from shading: a new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43.

[41] Durou, J.-D., Aujol, J.-F., and Courteille, F. (2009). Integrating the normal field of a surface in the presence of discontinuities. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 261–273.

[42] Eigen, D., Puhrsch, C., and Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, pages 2366–2374.

[43] Endres, F., Hess, J., Sturm, J., Cremers, D., and Burgard, W. (2014). 3-D mapping with an RGB-D camera. *IEEE Transactions on Robotics*, 30(1):177–187.

[44] Engel, J., Schöps, T., and Cremers, D. (2014). LSD-SLAM: Large-scale direct monocular SLAM. In *13th European Conference on Computer Vision (ECCV)*, pages 834–849.

[45] Favaro, P., Mennucci, A., and Soatto, S. (2003). Observing shape from defocused images. *International Journal of Computer Vision (IJCV)*, 52(1):25–43.

[46] Favaro, P. and Papadhimitri, T. (2012). A closed-form solution to uncalibrated photometric stereo via diffuse maxima. In *2012 IEEE Conference onComputer Vision and Pattern Recognition (CVPR)*, pages 821–828.

[47] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

[48] Frankot, R. T. and Chellappa, R. (1988). A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 10(4):439–451.

[49] Furukawa, Y. and Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(8):1362–1376.

[50] Garg, R., BG, V. K., Carneiro, G., and Reid, I. (2016). Unsupervised cnn for single view depth estimation: Geometry to the rescue. In *14th European Conference on Computer Vision (ECCV)*, pages 740–756.

[51] Georghiades, A. S. (2003). Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *International Conference on Computer Vision (ICCV)*, pages 816–825.

[52] Ghosh, A., Heidrich, W., Achutha, S., and O'Toole, M. (2010). A basis illumination approach to BRDF measurement. *International Journal of Computer Vision (IJCV)*, 90(2):183–197.

[53] Goldman, D. B., Curless, B., Hertzmann, A., and Seitz, S. M. (2010). Shape and spatially-varying BRDFs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(6):1060–1071.

[54] Haefner, B., Quéau, Y., Möllenhoff, T., and Cremers, D. (2018). Fight ill-posedness with ill-posedness: Single-shot variational depth super-resolution from shading. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 164–174.

[55] Hager, G. D. and Belhumeur, P. N. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(10):1025–1039.

[56] Harltey, A. and Zisserman, A. (2006). *Multiple view geometry in computer vision (2. ed.)*. Cambridge University Press.

[57] Haussecker, H. W. and Fleet, D. J. (2001). Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(6):661–673.

[58] Hernández, C., Vogiatzis, G., Brostow, G. J., Stenger, B., and Cipolla, R. (2007). Non-rigid Photometric Stereo with Colored Lights. In *2007 IEEE International Conference on Computer Vision (ICCV)*, pages 1–8.

[59] Hernández, C., Vogiatzis, G., and Cipolla, R. (2008). Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 30(3):548–554.

[60] Hernández, C., Vogiatzis, G., and Cipolla, R. (2011). Overcoming shadows in 3-source photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(2):419–426.

[61] Hertzmann, A. and Seitz, S. (2005). Example-based photometric stereo: shape reconstruction with general, varying brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(8):1254–1264.

[62] Higo, T., Matsushita, Y., Joshi, N., and Ikeuchi, K. (2009). A hand-held photometric stereo camera for 3-D modeling. In *2009 IEEE International Conference on Computer Vision (ICCV)*, pages 1234–1241.

[63] Ho, J., Lim, J., Yang, M.-H., and Kriegman, D. (2006). Integrating surface normal vectors using fast marching method. In *9th European Conference on Computer Vision (ECCV)*, pages 239–250.

[64] Hold-Geoffroy, Y., Zhang, J., Gotardo, P. F. U., and Lalonde, J.-F. (2015). *X*-hour Outdoor Photometric Stereo. In *3DV*.

[65] Hong, J. H. (2018). *Widening the basin of convergence for the bundle adjustment type of problems in computer vision*. PhD thesis, Department of Engineering,University of Cambridge,UK.

[66] Horn, B. and Brooks, M. (1986). The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing (CVGIP)*, 33(2):174–208.

[67] Horn, B. K. P. (1970). *Shape from Shading: a Method for Obtaining the Shape of a Smooth Opaque Object from One View.* PhD thesis, Department of Electrical Engieerining and Computer Science, Massachusetts Institute of Technology.

[68] Horn, B. K. P. (1975). Obtaining shape from shading information. *The Psychology of Computer Vision, Winston, P. H. (Ed.)*, pages 115–155.

[69] Hullin, M. B., Hanika, J., Ajdin, B., Seidel, H.-P., Kautz, J., and Lensch, H. (2010). Acquisition and analysis of bispectral bidirectional reflectance and reradiation distribution functions. *ACM Transactions on Graphics (TOG)*, 29(4).

[70] Hung, C.-H., Wu, T.-P., Matsushita, Y., Xu, L., Jia, J., and Tang, C.-K. (2015). Photometric stereo in the wild. In *2015 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 302–309.

[71] Hyeongwoo, K., Hailin, J., Hadap, S., and Inso, K. (2013). Specular reflection separation using dark channel prior. In *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1460–1467.

[72] Ikehata, S., Wipf, D., Matsushita, Y., and Aizawa, K. (2012). Robust photometric stereo using sparse regression. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 318–325.

[73] Ikeuchi, K. (1981). Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 3(6):661–669.

[74] Iwahori, Y., Sugie, H., and Ishii, N. (1990). Reconstructing shape from shading images under point light source illumination. In *International Conference on Pattern Recognition (ICPR)*, pages 83–87.

[75] Jacobs, D. W., Belhumeur, P. N., and Basri, R. (1998). Comparing images under variable illumination. In *1998 IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 610–617.

[76] Jin, H., Cremers, D., Wang, D., Prados, E., Yezzi, A., and Soatto, S. (2008). 3-D reconstruction of shaded objects from multiple images under unknown illumination. *International Journal of Computer Vision (IJCV)*, 76(3):245–256.

[77] Jin, H., Cremers, D., Yezzi, A. J., and Soatto, S. (2004). Shedding light on stereoscopic segmentation. In *2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–I.

[78] Jin, H., Favaro, P., and Soatto, S. (2001). Real-time feature tracking and outlier rejection with changes in illumination. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV*.

[79] Jin, H., Soatto, S., and Yezzi, A. J. (2003). Multi-view stereo beyond Lambert. In *2003 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–I.

[80] Johnson, M. K. and Adelson, E. H. (2009). Retrographic sensing for the measurement of surface texture and shape. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1070–1077.

[81] Johnson, M. K., Cole, F., Raj, A., and Adelson, E. H. (2011). Microgeometry capture using an elastomeric sensor. In *Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*.

[82] Joshi, N. and Kriegman, D. J. (2007). Shape from varying illumination and viewpoint. In *2007 IEEE International Conference on Computer Vision (ICCV)*, pages 1–7.

[83] Jung, J., Lee, J. Y., and Kweon, I. S. (2015). One-day outdoor photometric stereo via skylight estimation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4521–4529.

[84] Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, pages 61–70.

[85] Kazhdan, M., Klein, A., Dalal, K., and Hoppe, H. (2007). Unconstrained isosurface extraction on arbitrary octrees. In *Eurographics Symposium on Geometry Processing*.

[86] Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., and Bry, A. (2017). End-to-end learning of geometry and context for deep stereo regression. In *IEEE International Conference on Computer Vision (ICCV)*.

[87] Kerl, C., Sturm, J., and Cremers, D. (2013). Dense visual slam for rgb-d cameras. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2100–2106.

[88] Kimmel, R. and Bruckstein, A. M. (1995). Tracking level sets by level sets: A method for solving the shape from shading problem. *CVIU*, 62(1):47–58.

[89] Klein, G. and Murray, D. (2007). Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE.

[90] Konstantinou, C., Biscontin, G., and Logothetis, F. (2019). Tensile strength of bio-cemented sands. Under Review for Journal of rock mechanics and rock engineering.

[91] Koshikawa, K. (1979). A polarimetric approach to shape understanding of glossy objects. In *Proceedings of the 6th international joint conference on Artificial intelligence-Volume 1*, pages 493–495. Morgan Kaufmann Publishers Inc.

[92] Kumar, J. B., Prabhakara, R., and Pushpa, H. (2013). Bio mineralisation of calcium carbonate by different bacterial strains and their application in concrete crack remediation. *International Journal of Advances in Engineering & Technology*, 6(1):202.

[93] Kurt, M. and Edwards, D. (2009). A survey of brdf models for computer graphics. *ACM SIGGRAPH computer graphics*, 43(2):4:1–4:7.

[94] Lafortune, E. P., Foo, S.-C., Torrance, K. E., and Greenberg, D. P. (1997). Non-linear approximation of reflectance functions. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 117–126. ACM Press/Addison-Wesley Publishing Co.

[95] Lagarde, S. (2011). Adopting a physically based shading model. Retrieved on 1/10/2018 from https://seblagarde.wordpress.com/2011/08/17/hello-world/#comment-2405.

[96] Lambert, J. H. (1760). *Photometria sive De mensura et gradibus luminis, colorum et umbrae*. Sumptibus viduae Eberhardi Klett, typis Christophori Petri Detleffsen.

[97] Lee, K. M. and Kuo, C.-C. (1996). Shape from photometric ratio and stereo. *Journal of Visual Communication and Image Representation*, 7(2):155 – 162.

[98] Lee, S. and Brady, M. (1991). Integrating stereo and photometric stereo to monitor the development of glaucoma. *Image and Vision Computing*, 9(1):39–44.

[99] Leonard, J. J. and Durrant-Whyte, H. F. (1991). Simultaneous map building and localization for an autonomous mobile robot. In *Proceedings IROS '91:IEEE/RSJ International Workshop on Intelligent Robots and Systems '91*, pages 1442–1447 vol.3.

[100] Lim, J., Ho, J., Yang, M., and Kriegman, D. J. (2005). Passive photometric stereo from motion. In *2005 IEEE International Conference on Computer Vision (ICCV)*, pages 1635–1642.

[101] Liu, C., Narasimhan, S. G., and Dubrawski, A. W. (2018). Near-light photometric stereo using circularly placed point light sources. In *Computational Photography (ICCP), 2018 IEEE International Conference on*, pages 1–10.

[102] Liu, L., Zhang, L., Xu, Y., Gotsman, C., and Gortler, S. J. (2008). A local/global approach to mesh parameterization. *Comput. Graph. Forum*.

[103] Logothetis, F., Mecca, R., and Cipolla, R. (2017). Semi-calibrated near field photometric stereo. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 3, page 8.

[104] Logothetis, F., Mecca, R., and Cipolla, R. (2018a). A differential volumetric approach to multi-view photometric stereo. *arXiv*, 1811.01984.

[105] Logothetis, F., Mecca, R., Quéau, Y., and Cipolla, R. (2016). Near-field photometric stereo in ambient light. In *British Machine Vision Conference (BMVC)*.

[106] Logothetis, F., Mecca, R., Sgallari, F., and Cipolla, R. (2018b). A differential approach to shape from polarisation: a level-set characterisation. *International Journal of Computer Vision (IJCV)*, pages 1–14.

[107] Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828):133.

[108] Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. In *ACM SIGGRAPH computer graphics*, volume 21, pages 163–169.

[109] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision (IJCV)*, 60(2):91–110.

[110] Luebke, D., Reddy, M., Cohen, J. D., Varshney, A., Watson, B., and Huebner, R. (2003). *Level of Detail for 3D Graphics*. Morgan Kaufmann.

[111] Maier, R., Kim, K., Cremers, D., Kautz, J., and Niessner, M. (2017). Intrinsic3D: High-quality 3D Reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In *2017 IEEE International Conference on Computer Vision (ICCV)*.

[112] Malladi, R., Sethian, J. A., and Vemuri, B. C. (1995). Shape modeling with front propagation: A level set approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(2):158–175.

[113] Marschner, S. R., Westin, S. H., Lafortune, E. P. F., Torrance, K. E., and Greenberg, D. P. (1999). Image-based BRDF measurement including human skin. In *Rendering Techniques' 99*, pages 131–144.

[114] Mecca, R. and Durou, J. D. (2011). Unambiguous photometric stereo using two images. In *Image Analysis and Processing - ICIAP 2011 - 16th International Conference, Ravenna, Italy, September 14-16*, pages 286–295.

[115] Mecca, R. and Falcone, M. (2013). Uniqueness and approximation of a photometric shape-from-shading model. *SIAM J. Imag. Sci.*, 6(1):616–659.

[116] Mecca, R., Logothetis, F., and Cipolla, R. (2017). A differential approach to shape from polarization. In *British Machine Vision Conference (BMVC)*.

[117] Mecca, R. and Quéau, Y. (2016). Unifying Diffuse and Specular Reflections for the Photometric Stereo Problem. In *Workshop on Applications of Computer Vision (WACV)*.

[118] Mecca, R., Quéau, Y., Logothetis, F., and Cipolla, R. (2016). A single lobe photometric stereo approach for heterogeneous material. *SIAM Journal on Imaging Sciences*, 9(4):1858–1888.

[119] Mecca, R., Rodolà, E., and Cremers, D. (2015). Realistic photometric stereo using partial differential irradiance equation ratios. *Computers & Graphics*, 51:8–16.

[120] Mecca, R., Tankus, A., Wetzler, A., and Bruckstein, A. (2014a). A direct differential approach to photometric stereo with perspective viewing. *SIAM Journal on Imaging Sciences*, 7(2):579–612.

[121] Mecca, R., Wetzler, A., Bruckstein, A., and Kimmel, R. (2014b). Near Field Photometric Stereo with Point Light Sources. *SIAM Journal on Imaging Sciences*, 7(4):2732–2770.

[122] Mecca, R., Wetzler, A., Kimmel, R., and Bruckstein, A. M. (2013). Direct shape recovery from photometric stereo with shadows. In *International Conference on 3D Vision (3DV) 2013*, pages 382–389.

[123] Menze, M. and Geiger, A. (2015). Object scene flow for autonomous vehicles. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3061–3070.

[124] Migita, T., Ogino, S., and Shakunaga, T. (2008). Direct bundle estimation for recovery of shape, reflectance property and light position. In *10th European Conference on Computer Vision (ECCV)*, pages 412–425.

[125] Moses, Y. and Shimshoni, I. (2009). 3d shape recovery of smooth surfaces: Dropping the fixed-viewpoint assumption. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(7):1310–1324.

[126] Murez, Z., Treibitz, T., Ramamoorthi, R., and Kriegman, D. (2015). Photometric stereo in a scattering medium. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3415–3423.

[127] Nayar, S. K. (1989). Sphereo: Determining depth using two specular spheres and a single camera. In *Optics, Illumination, and Image Sensing for Machine Vision III*, volume 1005, pages 245–254.

[128] Nayar, S. K., Ikeuchi, K., and Kanade, T. (1990). Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(4):418–431.

[129] Nehab, D., Rusinkiewicz, S., Davis, J., and Ramamoorthi, R. (2005). Efficiently combining positions and normals for precise 3d geometry. *ACM transactions on graphics (TOG)*, 24(3):536–543.

[130] Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S., and Fitzgibbon, A. (2011). Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE international symposium on Mixed and augmented reality (ISMAR)*, pages 127–136.

[131] NextEngine (2010). *3D Scanner Ultra HD*. NextEngine. http://www.nextengine.com/products.

[132] Ngan, A., Durand, F., and Matusik, W. (2005). Experimental analysis of BRDF models. In *Proceedings of the Sixteenth Eurographics conference on Rendering Techniques*, pages 117–126.

[133] Nicodemus, F. E. (1965). Directional reflectance and emissivity of an opaque surface. *Applied optics*, 4(7):767–775.

[134] Nießner, M., Zollhöfer, M., Izadi, S., and Stamminger, M. (2013). Real-time 3d reconstruction at scale using voxel hashing. *ACM Transactions on Graphics (ToG)*, 32(6):169.

[135] Ochs, P., Chen, Y., Brox, T., and Pock, T. (2014). ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419.

[136] Okatani, T. and Deguchi, K. (2012). Optimal integration of photometric and geometric surface measurements using inaccurate reflectance/illumination knowledge. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 254–261.

[137] Oliensis, J. (1991). Uniqueness in shape from shading. *International Journal of Computer Vision (IJCV)*, 6(2):75–104.

[138] Onn, R. and Bruckstein, A. M. (1990). Integrability disambiguates surface recovery in two-image photometric stereo. *International Journal of Computer Vision (IJCV)*, 5(1):105–113.

[139] Or-el, R., Rosman, G., Wetzler, A., Kimmel, R., and Bruckstein, A. M. (2015). RGBD-Fusion: Real-Time High Precision Depth Recovery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA.

[140] Osher, S. and Fedkiw, R. (2006). *Level Set Methods and Dynamic Implicit Surfaces*. Applied Mathematical Sciences. Springer New York.

[141] OSRAM (2009). *Platinum DRAGON LED*. OSRAM - Opto Semicondutors.

[142] Papadhimitri, T. and Favaro, P. (2013). A New Perspective on Uncalibrated Photometric Stereo. In *2013 IEEE Conference onComputer Vision and Pattern Recognition (CVPR)*, pages 1474–1481.

[143] Papadhimitri, T. and Favaro, P. (2014). Uncalibrated Near-Light Photometric Stereo. In *British Machine Vision Conference (BMVC)*.

[144] Park, J., Sinha, S. N., Matsushita, Y., Tai, Y. W., and Kweon, I. S. (2017). Robust multiview photometric stereo using planar mesh parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 39(8):1591–1604.

[145] Pham-Gia, T., Turkkan, N., and Marchand, E. (2006). Density of the ratio of two normal random variables and applications. *Communications in Statistics - Theory and Methods*, 35(9):1569–1591.

[146] Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317.

[147] Ping-Sing, T. and Shah, M. (1994). Shape from shading using linear approximation. *Image and Vision computing*, 12(8):487–498.

[148] Prados, E. and Faugeras, O. D. (2005). Shape from shading: A well-posed problem? In *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 870–877.

[149] Quéau, Y., Durix, B., Wu, T., Cremers, D., Lauze, F., and Durou, J.-D. (2018). Led-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision (JMIV)*, 60(3):313–340.

[150] Quéau, Y. and Durou, J.-D. (2015). Edge-preserving integration of a normal field: Weighted least squares, TV and L1 approaches. In *SSVM*.

[151] Quéau, Y., Lauze, F., and Durou, J.-D. (2015). A L1-TV Algorithm for Robust Perspective Photometric Stereo with Spatially-Varying Lightings. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 498–510.

[152] Quéau, Y., Mecca, R., and Durou, J.-D. (2016). Unbiased photometric stereo for colored surfaces: A variational approach. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4359–4368.

[153] Quéau, Y., Wu, T., and Cremers, D. (2017a). Semi-calibrated near-light photometric stereo. In *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 656–668.

[154] Quéau, Y., Wu, T., Lauze, F., Durou, J.-D., and Cremers, D. (2017b). A non-convex variational approach to photometric stereo under inaccurate lighting. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 3, pages 14–15.

[155] Radow, G., Hoeltgen, L., Quéau, Y., and Breuß, M. (2018). Optimisation of classic photometric stereo by non-convex variational minimisation. *Journal of Mathematical Imaging and Vision (JMIV)*, pages 1–22.

[156] Rindfleisch, T. (1966). Photometric method for lunar topography. *Photometric Engineering*, 32(2):262–277.

[157] Sabzevari, R., Bue, A. D., and Murino, V. (2012). Multi-view photometric stereo using semi-isometric mappings. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*.

[158] Santo, H., Samejima, M., Sugano, Y., Shi, B., and Matsushita, Y. (2017). Deep photometric stereo network. In *ICCV Workshops*.

[159] Sato, I., Okabe, T., and Sato, Y. (2012). Bispectral photometric stereo based on fluorescence. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 270–277.

[160] Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision (IJCV)*, 47(1-3):7–42.

[161] Schlick, C. (1994). An inexpensive BRDF model for physically-based rendering. In *Computer graphics forum*, volume 13, pages 233–246. Wiley Online Library.

[162] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 519–528.

[163] Selesnick, I. (2009). A derivation of the soft-thresholding function. Polytechnic Institute of New York University.

[164] Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595.

[165] Sheffer, A., Praun, E., and Rose, K. (2006). Mesh parameterization methods and their applications. *Foundations and Trends in Computer Graphics and Vision*.

[166] Shen, F., Sunkavalli, K., Bonneel, N., Rusinkiewicz, S., Pfister, H., and Tong, X. (2014). Time-Lapse Photometric Stereo and Applications. *Computer Graphics Forum*, 33(7):359–367.

[167] Shi, B., Matsushita, Y., Wei, Y., Xu, C., and Tan, P. (2010). Self-calibrating photometric stereo. In *2010 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, pages 1118–1125.

[168] Simchony, T., Chellappa, R., and Shao, M. (1990). Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(5):435–446.

[169] Smith, W. and Fang, F. (2016). Height from photometric ratio with model-based light source selection. *Computer Vision and Image Understanding*, 145:128–138.

[170] Smith, W. A., Ramamoorthi, R., and Tozza, S. (2016). Linear depth estimation from an uncalibrated, monocular polarisation image. In *14th European Conference on Computer Vision (ECCV)*, pages 109–125.

[171] Sunkavalli, K., Romeiro, F., Matusik, W., Zickler, T., and Pfister, H. (2008). What do color changes reveal about an outdoor scene? In *2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8.

[172] Sunkavalli, K., Zickler, T. E., and Pfister, H. (2010). Visibility subspaces: Uncalibrated photometric stereo with shadows. In *11th European Conference on Computer Vision (ECCV)*, pages 251–264.

[173] Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Texts in Computer Science. Springer London.

[174] Tankus, A. and Kiryati, N. (2005). Photometric stereo under perspective projection. In *2005 IEEE International Conference on Computer Vision (ICCV)*, pages 611–616.

[175] Tikhonov, A. and Arsenin, V. (1979). Solutions of ill-posed problems. *SIAM Review*, 21(2):266.

[176] Torrance, K. E. and Sparrow, E. M. (1967). Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America (JOSA)*, 57(9):1105–1114.

[177] Tozza, S., Mecca, R., Duocastella, M., and Bue, A. D. (2016). Direct differential photometric stereo shape recovery of diffuse and specular surfaces. *Journal of Mathematical Imaging and Vision (JMIV)*, 56(1):57–76.

[178] Treibitz, T., Murez, Z., Mitchell, B. G., and Kriegman, D. (2012). Shape from fluorescence. In *12th European Conference on Computer Vision (ECCV)*, pages 292–306.

[179] Treibitz, T. and Schechner, Y. Y. (2009). Active polarization descattering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(3):385–399.

[180] Treuille, A., Hertzmann, A., and Seitz, S. M. (2004). Example-based stereo with general BRDFs. In *8th European Conference on Computer Vision (ECCV)*, pages 457–469.

[181] Tsiotsios, C., Angelopoulou, M. E., Kim, T., and Davison, A. J. (2014). Backscatter compensated photometric stereo with 3 sources. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2251–2258.

[182] Van Diggelen, J. (1951). A photometric investigation of the slopes and the heights of the ranges of hills in the maria of the moon. *Bulletin of the Astronomical Institutes of the Netherlands*, 11:283.

[183] Vlasic, D., Peers, P., Baran, I., Debevec, P. E., Popovic, J., Rusinkiewicz, S., and Matusik, W. (2009). Dynamic shape capture using multi-view photometric stereo. *ACM Transactions on Graphics*, 28(5).

[184] Vogiatzis, G., Torr, P. H. S., and Cipolla, R. (2005). Multi-view stereo via volumetric graph-cuts. In *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 391–398.

[185] Wang, Y., Lee, C. M., Cheong, L., and Toh, K. (2015). Practical matrix completion and corruption recovery using proximal alternating robust subspace minimization. *International Journal of Computer Vision (IJCV)*, 111(3):315–344.

[186] Ward, G. J. (1992). Measuring and modeling anisotropic reflection. *ACM SIGGRAPH Computer Graphics*, 26(2):265–272.

[187] Weber, M., Blake, A., and Cipolla, R. (2002). Towards a complete dense geometric and photometric reconstruction under varying pose and illumination. In *British Machine Vision Conference (BMVC)*.

[188] Westin, S. H., Arvo, J. R., and Torrance, K. E. (1992). Predicting reflectance functions from complex surfaces. *Computer Graphics*, 26(2):255–264.

[189] Whelan, T., Salas-Moreno, R. F., Glocker, B., Davison, A. J., and Leutenegger, S. (2016). Elasticfusion: Real-time dense slam and light source estimation. *The International Journal of Robotics Research*, 35(14):1697–1716.

[190] Wolff, L. B. (1994). Relative brightness of specular and diffuse reflection. *Optical Engineering*, 33(1):285–294.

[191] Wolff, L. B. and Angelopoulou, E. (1994). 3-D stereo using photometric ratios. In *3rd European Conference on Computer Vision (ECCV)*, pages 247–258.

[192] Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):134–144.

[193] Wu, C. (2013). Towards linear-time incremental structure from motion. In *3D Vision-3DV 2013, 2013 International conference on*, pages 127–134.

[194] Wu, C., Agarwal, S., Curless, B., and Seitz, S. M. (2011a). Multicore bundle adjustment. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3057–3064.

[195] Wu, C., Liu, Y., Dai, Q., and Wilburn, B. (2011b). Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination. *IEEE transactions on visualization and computer graphics*, 17(8):1082–1095.

[196] Wu, C., Narasimhan, S. G., and Jaramaz, B. (2010a). A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes. *International Journal of Computer Vision (IJCV)*, 86(2-3):211–228.

[197] Wu, C., Varanasi, K., Liu, Y., Seidel, H., and Theobalt, C. (2011c). Shading-based dynamic shape refinement from multi-view video under general illumination. In *2011 IEEE International Conference on Computer Vision (ICCV)*, pages 1108–1115.

[198] Wu, C., Zollhöfer, M., Nieß ner, M., Stamminger, M., Izadi, S., and Theobalt, C. (2014). Real-time shading-based refinement for consumer depth cameras. *ACM Transactions on Graphics (TOG)*, 33(6):200.

[199] Wu, L., Ganesh, A., Shi, B., Matsushita, Y., Wang, Y., and Ma, Y. (2010b). Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision (ACCV)*, pages 703–717.

[200] Wu, T.-P., Tang, K.-L., Tang, C.-K., and Wong, T.-T. (2006). Dense photometric stereo: A markov random field approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28(11):1830–1846.

[201] Xie, W., Dai, C., and Wang, C. C. L. (2015). Photometric stereo with near point lighting: A solution by mesh deformation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4585–4593.

[202] Yang, Q. and Ahuja, N. (2012). Surface reflectance and normal estimation from photometric stereo. *Computer Vision and Image Understanding*, 116(7):793–802.

[203] Ylimäki, M., Kannala, J., and Heikkilä, J. (2017). Robust and practical depth map fusion for time-of-flight cameras. In *Image Analysis*, pages 122–134.

[204] Yoshiyasu, Y. and Yamazaki, N. (2011). Topology-adaptive multi-view photometric stereo. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1001–1008.

[205] Yuille, A. L., Snow, D., Epstein, R., and Belhumeur, P. N. (1999). Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability. *International Journal of Computer Vision (IJCV)*, 35(3):203–222.

[206] Zeng, G., Paris, S., Quan, L., and Sillion, F. (2005). Progressive surface reconstruction from images using a local prior. In *2005 IEEE International Conference on Computer Vision (ICCV)*, pages 1230–1237.

[207] Zhang, L., Curless, B., Hertzmann, A., and Seitz, S. M. (2003). Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multiview stereo. In *2003 IEEE International Conference on Computer Vision (ICCV)*, pages 618–625.

[208] Zhang, R., Tsai, P.-S., Cryer, J. E., and Shah, M. (1999). Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(8):690–706.

[209] Zhou, Z., Wu, Z., and Tan, P. (2013). Multi-view photometric stereo with spatially varying isotropic materials. In *2013 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 1482–1489.

[210] Zollhöfer, M., Dai, A., Innmann, M., Wu, C., Stamminger, M., Theobalt, C., and Nießner, M. (2015). Shading-based refinement on volumetric signed distance functions. *ACM Transactions on Graphics (TOG)*, 34(4):96.

Appendix A

Proofs

A.1 Unified equation

This section includes the proofs about the properties of the unified equation introduced in Section 3.5 and are included there for completeness.

Theorem 3.6.1 Proof We prove it by contradiction by assuming that there exists a point $(\widetilde{u}, \widetilde{v}, \widetilde{z}(\widetilde{u}, \widetilde{v}))$ such that

$$(a_k i_j)^c \left(\overline{\boldsymbol{W}}_k^1 - \frac{\widetilde{u}}{f}\overline{\boldsymbol{W}}_k^3\right) - (a_j i_k)^c \left(\overline{\boldsymbol{W}}_j^1 - \frac{\widetilde{u}}{f}\overline{\boldsymbol{W}}_j^3\right) = 0, \tag{A.1}$$

$$(a_k i_j)^c \left(\overline{\boldsymbol{W}}_k^2 - \frac{\widetilde{v}}{f}\overline{\boldsymbol{W}}_k^3\right) - (a_j i_k)^c \left(\overline{\boldsymbol{W}}_j^2 - \frac{\widetilde{v}}{f}\overline{\boldsymbol{W}}_j^3\right) = 0, \tag{A.2}$$

By considering the image irradiance Equation 3.16, we get $i_j^c = (\rho a_j)^c(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j)$ and $i_k^c = (\rho a_k)^c(\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_k)$ thus substituting in (A.1) and (A.2), we get the following equations

$$(a_j a_k \rho)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j \left(\overline{\boldsymbol{W}}_k^1 - \frac{\widetilde{u}}{f}\overline{\boldsymbol{W}}_k^3\right) - (a_j a_k \rho)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_k \left(\overline{\boldsymbol{W}}_j^1 - \frac{\widetilde{u}}{f}\overline{\boldsymbol{W}}_j^3\right) = 0, \tag{A.3}$$

$$(a_j a_k \rho)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j \left(\overline{\boldsymbol{W}}_k^2 - \frac{\widetilde{v}}{f}\overline{\boldsymbol{W}}_k^3\right) - (a_j a_k \rho)^c \overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_k \left(\overline{\boldsymbol{W}}_j^2 - \frac{\widetilde{v}}{f}\overline{\boldsymbol{W}}_j^3\right) = 0 \tag{A.4}$$

which can be simplified as follows

$$\boldsymbol{N} \cdot \left(\overline{\boldsymbol{W}}_j \left(\overline{\boldsymbol{W}}_k^1 - \frac{\widetilde{u}}{f}\overline{\boldsymbol{W}}_k^3\right) - \overline{\boldsymbol{W}}_k \left(\overline{\boldsymbol{W}}_j^1 - \frac{\widetilde{u}}{f}\overline{\boldsymbol{W}}_j^3\right)\right) = 0, \tag{A.5}$$

$$\boldsymbol{N} \cdot \left(\overline{\boldsymbol{W}}_j \left(\overline{\boldsymbol{W}}_k^2 - \frac{\widetilde{v}}{f}\overline{\boldsymbol{W}}_k^3\right) - \overline{\boldsymbol{W}}_k \left(\overline{\boldsymbol{W}}_j^2 - \frac{\widetilde{v}}{f}\overline{\boldsymbol{W}}_j^3\right)\right) = 0. \tag{A.6}$$

If we solve the system made by the equations (A.5) and (A.6) with respect to $\widetilde{z}_u$ and $\widetilde{z}_v$ (using the normal parameterisation from Equation 3.8), we get the following solution

$$\widetilde{z}_u = \frac{(\overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^3 - \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^2)(f+\widetilde{z})}{f\overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^1 - f\overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_k^2 + \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^2\widetilde{u} - \overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^3\widetilde{u} - \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^1\widetilde{v} + \overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_k^3\widetilde{v}} \tag{A.7}$$

$$\widetilde{z}_v = - \frac{(\overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_k^3 - \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^1)(f+\widetilde{z})}{f\overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^1 - f\overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_k^2 + \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^2\widetilde{u} - \overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^3\widetilde{u} - \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^1\widetilde{v} + \overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_k^3\widetilde{v}} \tag{A.8}$$

and once we compute the image values substituting $\widetilde{z}_u$ and $\widetilde{z}_v$ into the irradiance Equation 3.16, we get $i_j = 0$ and $i_k = 0$ which can never be true for points in the image domain with $\rho > 0$ and not in a shadow region and would invalidate the ratio method anyway.

Theorem 3.6.2 Proof Again, we prove by contradiction by assuming that there exists a point $(\widetilde{u}, \widetilde{v}, \widetilde{z}(\widetilde{u}, \widetilde{v}))$ violating the linear independence. Then:

$$\boldsymbol{b}_{jk}(\widetilde{u}, \widetilde{v}, \widetilde{z}) \cdot \boldsymbol{b}_{jq}(\widetilde{u}, \widetilde{v}, \widetilde{z}) = \pm|\boldsymbol{b}_{jk}(\widetilde{u}, \widetilde{v}, \widetilde{z})|\ |\boldsymbol{b}_{jq}(\widetilde{u}, \widetilde{v}, \widetilde{z})|. \tag{A.9}$$

To improve readability we omit the dependence on $(\widetilde{u}, \widetilde{v}, \widetilde{z})$. Now, by squaring both sides we have:

$$\left(b_{jk}^1 b_{jq}^1 + b_{jq}^2 b_{jq}^2\right)^2 = \left(\left(b_{jk}^1\right)^2 + \left(b_{jk}^2\right)^2\right)\left(\left(b_{jq}^1\right)^2 + \left(b_{jq}^2\right)^2\right). \tag{A.10}$$

Using the terms from Equation 3.19, replacing the respective irradiance equation in Equation 3.17 for i_j, i_k and i_q, eliminating the not null quantities like the albedo and the light attenuations, this can be rewritten as follows:

$$\begin{aligned}
&\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_q(f\overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_k^2 - f\overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^1 + \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^2\widetilde{u} - \overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^3\widetilde{u} + \overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_k^3\widetilde{v} - \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_k^1\widetilde{v}) + \\
&\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_k(f\overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_q^1 - f\overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_q^2 - \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_q^2\widetilde{u} + \overline{\boldsymbol{W}}_j^2\overline{\boldsymbol{W}}_q^3\widetilde{u} - \overline{\boldsymbol{W}}_j^1\overline{\boldsymbol{W}}_q^3\widetilde{v} + \overline{\boldsymbol{W}}_j^3\overline{\boldsymbol{W}}_q^1\widetilde{v}) + \\
&\overline{\boldsymbol{N}} \cdot \overline{\boldsymbol{W}}_j(f\overline{\boldsymbol{W}}_k^1\overline{\boldsymbol{W}}_q^2 - f\overline{\boldsymbol{W}}_k^2\overline{\boldsymbol{W}}_q^1 + \overline{\boldsymbol{W}}_k^3\overline{\boldsymbol{W}}_q^2\widetilde{u} - \overline{\boldsymbol{W}}_k^2\overline{\boldsymbol{W}}_q^3\widetilde{u} + \overline{\boldsymbol{W}}_k^1\overline{\boldsymbol{W}}_q^3\widetilde{v} - \overline{\boldsymbol{W}}_k^3\overline{\boldsymbol{W}}_q^1\widetilde{v}) = 0.
\end{aligned}$$

After noticing that all the terms in the previous equation are divided by the same not null quantities as $|\boldsymbol{N}|$ and $|\boldsymbol{W}_i|$, we can consider the not-normalized vectors (i.e., we can remove the bar on top of each factor). Then, substituting the normal parameterisation from Equation 3.8, we get

$$\boldsymbol{W}_j^1\boldsymbol{W}_k^3\boldsymbol{W}_q^2 - \boldsymbol{W}_j^3\boldsymbol{W}_k^1\boldsymbol{W}_q^2 - \boldsymbol{W}_j^1\boldsymbol{W}_k^2\boldsymbol{W}_q^3 + \boldsymbol{W}_j^2\boldsymbol{W}_k^1\boldsymbol{W}_q^3 + \boldsymbol{W}_j^3\boldsymbol{W}_k^2\boldsymbol{W}_q^1 - \boldsymbol{W}_j^2\boldsymbol{W}_k^3\boldsymbol{W}_q^1 = 0 \tag{A.11}$$

Equation A.11 is the scalar triple product of vectors $\boldsymbol{W}_j, \boldsymbol{W}_k$ and $\boldsymbol{W}_q$. Having a 0 scalar triple product implies that the vectors are not linearly independent and thus $3 > \text{rank}([\mathbf{W}_j, \mathbf{W}_k, \mathbf{W}_q])$ which is a contradiction.